Designing, measuring and modelling the impact of the hands-on particle physics learning laboratory S'Cool LAB at CERN

Effects of student and laboratory characteristics on high-school students' cognitive and affective outcomes

Dissertation

Julia Woithe

Vom Fachbereich Physik der Technischen Universität Kaiserslautern zur Verleihung des akademischen Grades „Doktor der Naturwissenschaften“ genehmigte Dissertation

durchgeführt am
CERN
in Genf, Schweiz

unter Anleitung von
Prof. Dr. Jochen Kuhn
Fachbereich Physik der Technischen Universität Kaiserslautern

Zweitgutachter: Prof. Dr. Andreas Müller
Fachbereich Physik der Universität Genf

Datum der wissenschaftlichen Aussprache: 10. Juli 2020

D 386

CERN-THESIS-2020-089
10/07/2020

Abstract

Although many students find science important, science subjects are among the least favourite school subjects, student numbers declined over the past decades, and women are still strongly underrepresented in science. In particular, students' decreasing enjoyment of and interest in school science, their low science-related self-beliefs, their stereotypic images of science and scientists, and thus, their low interest in science and technology careers are worrying. However, today's students will need scientific literacy and scientific reasoning skills to face the challenges of the 21st century, such as the revolution of job markets due to machine learning and robotics.

To counteract this troubling trend, researchers call for a) more practical work, b) science learning that is relevant for students and takes their interest into account, c) activities fostering students' self-beliefs in science, in particular for female students, and d) more contact with scientists. Here, out-of-school learning opportunities are ideal complements to formal science education. In particular, out-of-school learning laboratories for high-school students hold a unique position within the informal learning sector because they provide hands-on activities, often guided by scientists, intended to foster students' interest in science. Furthermore, these labs provide authentic learning environments, which are often related to modern science topics, and research equipment that is usually not available in schools.

Whereas participation in hands-on workshops in these laboratories has an immense potential to foster both students' affects towards science and their knowledge and understanding, previous studies indicated that the high amount of novelty in these unique settings is not entirely beneficial. In particular, high levels of novelty can hinder learning. Here, student characteristics, such as their prior interest and experiences, and students' perceptions of characteristics of out-of-school learning labs, such as the organisation, cognitive load or support by volunteering scientists, are expected to influence the educational effectiveness of hands-on sessions in these settings. Hence, it is crucial to assess these novelty factors together with the desired educational outcomes of the learning activities to coherently understand the 'mechanism of action' of out-of-school learning labs and thereby optimise their educational potential.

In the framework of this PhD project, S'Cool LAB, a new out-of-school hands-on learning laboratory for high-school students was developed at CERN, Geneva, Switzerland. Particle physics learning activities were designed to help students understand the physics and technologies of the largest particle physics laboratory in the world, while also fostering their interest in physics, self-beliefs and scientific curiosity. To maximise the educational potential of this unique opportunity at CERN, the development of learning activities was based on documented students' conceptions and took recent research results about interest, curiosity, cognitive activation and cognitive load into account.

To analyse the educational effectiveness of half-day hands-on sessions in S'Cool LAB, and to understand the relevance of student and laboratory characteristics, more than 500 students from 15 different countries took part in a quantitative study employing a single group, pre- and post-test research design. In particular, different types of novelty factors were assessed and used to model S'Cool LAB's educational outcomes in both single-level and multi-level linear models.

The results of the S'Cool LAB evaluation study confirmed high educational effectiveness of its hands-on sessions. Despite the very short intervention time of only 4.5 hours, participation in a half-day hands-on session in S'Cool LAB led to medium-sized effects on students' physics interest ($d_C = 0.6$) and their conceptual understanding ($d_C = 0.7$), as well as a small effect on students' physics self-beliefs ($d_C = 0.4$). Here, girls benefited more

from the hands-on sessions in S'Cool LAB with respect to physics interest ($d_C = 0.8$) and self-beliefs ($d_{C,} = 0.5$) than did boys ($d_C = 0.5$ & 0.3). Consequently, the initial gender gap with girls reporting slightly lower dispositional physics interest ($d_C = 0.3$) and physics self-beliefs ($d_C = 0.4$) was closed with respect to their situational interest and self-beliefs.

In summary, S'Cool LAB outperforms other hands-on out-of-school learning labs by triggering very high situational interest and self-beliefs even for students whose dispositional interest and physics self-beliefs were already very high. Furthermore, S'Cool LAB demonstrates that out-of-school learning labs can foster students' conceptual understanding significantly without compromising the positive impact on other variables, such as interest and self-beliefs. When comparing students' epistemic curiosity towards particle physics before and after having taken part in hands-on sessions in S'Cool LAB, no significant differences were identified. However, the level of students' curiosity towards particle physics was already very high in the pre-visit questionnaire. Indeed, when comparing these curiosity levels with other topics, particle physics stands out as an extremely promising topic to both accommodate and maintain exceptionally high levels of epistemic curiosity. Students' perception of support by the learning environment and by the educators in S'Cool LAB were crucial elements of their novelty experience and were associated with especially high affective outcomes. In particular, students highly appreciated the volunteering CERN scientists and perceived their support as very valuable. Furthermore, students' perception of their cognitive preparedness and their cognitive load were vital with respect to the cognitive effects of S'Cool LAB. Overall, students' perception of laboratory characteristics accounted for twice the amount of variance in the educational outcomes than did student characteristics. Although the results of single- and multi-level modelling were similar, ignoring the hierarchical data structure was associated with a substantial overestimation of the effective sample size up to a factor of 11. Indeed, students' educational outcomes differed significantly among different groups of students because the groups showed major differences, for example, in their dispositional interest and their prior conceptual understanding.

In the framework of this PhD project, S'Cool LAB was successfully established as CERN's physics education research facility and students taking part in hands-on sessions also contributed to expanding our current understanding of students' conceptions about radiation and radioactivity. Here, the respective research results suggest that students' conceptual understanding of radiation and radioactivity has not changed much over the past decades. Instead, most of the previously documented conceptions were reproduced within this study. The results of a factor analysis suggest that, on the one hand, students argue rather consistently within certain areas, such as the macroscopic perspective on half-life. On the other hand, the results of a cluster analysis suggest that students' conceptual understanding of one area is independent of their understanding of a different area. Most strikingly, misconceptions seem to manifest themselves independently of students' (prior) knowledge. In summary, the results of the students' conceptions study confirm that it is possible and worthwhile to assess students' conceptual understanding of radiation and radioactivity quantitatively within the framework of out-of-school learning labs.

In summary, the S'Cool LAB studies show that the topic of particle physics provides a promising opportunity to introduce high-school students to modern physics by taking their interest and curiosity into account without compromising their conceptual understanding. Here, the high educational effectiveness of S'Cool LAB as a unique out-of-school learning lab confirms its successful development, more than justifies the required resources, and provides highly promising insights into the mechanism of action of out-of-school learning.

Zusammenfassung

Viele Jugendliche halten Naturwissenschaften für wichtig. Dennoch gehören naturwissenschaftliche Fächer zu den unbeliebtesten Schulfächern, Studierendenzahlen sind seit Jahrzehnten rückläufig und Frauen sind in den Naturwissenschaften nach wie vor stark unterrepräsentiert. Besorgniserregend ist dabei insbesondere, dass Jugendliche im Laufe ihrer Schulzeit die Freude und das Interesse an wissenschaftlichen Schulfächern weitestgehend verlieren, und sie zudem ein geringes Selbstkonzept und ein von Stereotypen geprägtes Bild von der Wissenschaft und von WissenschaftlerInnen zeigen. Dementsprechend ist ihr Interesse an naturwissenschaftlichen und technischen Berufen gering. Insbesondere in der heutigen Zeit sind jedoch wissenschaftliche Kenntnisse und Fähigkeiten, insbesondere das wissenschaftliche Denken, unabdingbar, um den Herausforderungen des 21. Jahrhunderts einschließlich der Revolution der Arbeitsmärkte durch Maschinenlernen und Robotik zu begegnen.

Um diesem besorgniserregenden Trend entgegenzuwirken, fordern ForscherInnen a) mehr Zeit für das Experimentieren, b) naturwissenschaftlichen Unterricht, der für Jugendliche relevant ist und deren Interesse berücksichtigt, c) Aktivitäten, die das Selbstkonzept der Schüler und insbesondere der Schülerinnen fördert, und d) mehr Kontakt zu WissenschaftlerInnen. Außerschulische Lernmöglichkeiten bieten hier eine ideale Ergänzung zum naturwissenschaftlichen Unterricht in der Schule. Dabei haben Schülerlabore eine einzigartige Position innerhalb des informellen Bildungssektors, da sie Experimentieraktivitäten anbieten, die häufig von WissenschaftlerInnen angeleitet werden und dadurch das Interesse der Jugendlichen an Naturwissenschaften fördern können. Darüber hinaus bieten Schülerlabore authentische Lernumgebungen, die oft moderne wissenschaftliche Themen sowie Forschungsequipment anbieten, welches in Schulen normalerweise nicht verfügbar ist.

Der Besuch von Schülerlaboren birgt ein immenses Potenzial, das konzeptuelle Verständnis von Jugendlichen sowie deren Einstellungen gegenüber Naturwissenschaften zu fördern. Allerdings deuteten frühere Studien darauf hin, dass das hohe Maß an ‚Novelty' in diesen speziellen Lernumgebungen nicht nur Vorteile hat, sondern das Lernen auch behindern kann. Es scheint, als könnten insbesondere gewisse Schülermerkmale (z. B. das Vorinteresse und Vorerfahrungen) sowie die Wahrnehmung von Schülerlabormerkmalen (z. B. die Organisation, kognitive Belastung oder Unterstützung durch betreuende WissenschaftlerInnen) die pädagogische Wirksamkeit von Schülerlaboren beeinflussen. Um die Wirkmechanismen von Schülerlaboren zu verstehen und deren Bildungspotenzial zu optimieren ist es daher essenziell, diese Novelty-Faktoren zusammen mit den erhofften Zielvariablen der Lernaktivitäten zu erheben.

Im Rahmen dieses Promotionsprojekts wurde am CERN (Genf, Schweiz) S'Cool LAB, ein neues Teilchenphysik-Schülerlabor aufgebaut. Insbesondere wurden Experimentierworkshops entwickelt, um Jugendlichen die Physik und Technologien des größten Teilchenphysiklabors der Welt verständlich zu machen und gleichzeitig ihr Interesse an Physik, ihr Physikselbstkonzept sowie ihre wissenschaftliche Neugier zu fördern. Um das Bildungspotential dieser einzigartigen Lerngelegenheit zu maximieren, wurden bei der Entwicklung der Lernaktivitäten dokumentierte Schülervorstellungen sowie aktuelle Forschungsergebnisse zu Interesse, Neugier, kognitiver Aktivierung und kognitiver Belastung berücksichtigt.

Um die pädagogische Wirksamkeit eines halbtägigen S'Cool LAB Besuchs und die Bedeutung von Schüler- und Labormerkmalen zu analysieren, nahmen mehr als 500 Jugendliche aus 15 verschiedenen Ländern an einer quantitativen Studie mit Prä-Post-Design teil. Dabei wurden verschiedene Arten von Novelty-Faktoren erfasst und als Prädiktoren sowohl in einfachen linearen Modellen als auch in Mehrebenen-Modellen untersucht. Die Ergebnisse der S'Cool LAB Evaluationsstudie bestätigten die hohe pädagogische Wirksamkeit dieses neuen Schülerlabors.

Trotz der sehr kurzen Interventionszeit von nur 4,5 Stunden führten Schülerlaborbesuche im S'Cool LAB zu mittelgroßen Effekten in Bezug auf das Physikinteresse der Jugendlichen ($d_C = 0{,}6$) und ihr konzeptuelles Verständnis ($d_C = 0{,}7$), sowie zu einem kleinen Effekt auf ihr Physikselbstkonzept ($d_C = 0{,}4$). Darüber hinaus profitierten Schülerinnen hinsichtlich des Physikinteresses ($d_C = 0.8$) und Selbstkonzepts ($d_{C,} = 0.5$) mehr vom Experimentieren im S'Cool LAB als Schüler ($d_C = 0.5$ & 0.3). Der ursprüngliche Gender Gap mit einem etwas geringeren dispositionalen Physikinteresse ($d_C = 0.3$) und dispositionalen Physikselbstkonzept ($d_C = 0.4$) der Schülerinnen konnte in Bezug auf das situationale Interesse und Selbstkonzept geschlossen werden.

Zusammenfassend lässt sich sagen, dass S'Cool LAB die Ergebnisse andere Schülerlabore übertrifft, indem es bei Jugendlichen zu sehr hohem situationalen Interesse und Selbstkonzept führt, selbst wenn die Jugendlichen bereits hohe dispositionale Eingangswerte aufwiesen. Darüber hinaus beweist S'Cool LAB, dass Schülerlabore sogar das konzeptuelle Verständnis von Jugendlichen erheblich fördern können, ohne dass dies andere Zielvariablen wie Interesse und Selbstkonzept beeinträchtigt. Der Schülerlaborbesuch führte zu keinen signifikanten Änderungen in Bezug auf die epistemische Neugier der Jugendlichen. Allerdings waren die Jugendlichen bereits vor ihrem Besuch höchst neugierig auf Teilchenphysik. Im Vergleich zu anderen Themen scheint es, dass Teilchenphysik ein äußerst vielversprechendes Thema ist, um eine außergewöhnlich hohe epistemische Neugierde von Jugendlichen zu berücksichtigen und aufrechtzuerhalten. Die Wahrnehmung der Lernatmosphäre sowie der TutorInnen im S'Cool LAB waren dabei entscheidende Novelty-Faktoren, die mit besonders hohen affektiven Zielvariablen verbunden waren. Die Jugendlichen nahmen die WissenschaftlerInnen sehr positiv wahr und empfanden deren Unterstützung als sehr wertvoll. Darüber hinaus war die Wahrnehmung der Jugendlichen hinsichtlich ihrer kognitiven Aktivierung und ihrer kognitiven Belastung von entscheidender Bedeutung für die kognitiven Effekte von S'Cool LAB. Insgesamt klärte die Wahrnehmung der Labormerkmale doppelt so viel Varianz der Zielvariablen auf als Schülermerkmale. Obwohl sich die Ergebnisse der einfachen und der Mehrebenen-Modelle ähnelten, überschätzten einfache lineare Modelle die effektive Stichprobe drastisch bis um das 11-fache, da diese die hierarchische Datenstruktur nicht berücksichtigten. In der Tat unterschieden sich die verschiedenen Schülergruppen in mehreren Zielvariablen erheblich, weil die Gruppen bereits große Unterschiede, zum Beispiel in ihrem dispositionellen Interesse und ihrem vorherigen konzeptuellen Verständnis, zeigten.

Im Rahmen dieses Promotionsprojekts wurde das Schülerlabor S'Cool LAB auch als Labor für fachdidaktische Forschung etabliert und TeilnehmerInnen unterstützten dabei auch Forschung zu Schülervorstellungen bezüglich Strahlung und Radioaktivität. Insbesondere zeigten die Ergebnisse, dass sich Schülervorstellungen zu Strahlung und Radioaktivität in den letzten Jahrzehnten nicht wesentlich geändert haben. Stattdessen wurden die meisten zuvor dokumentierten Schülervorstellungen in dieser Studie reproduziert. Die Ergebnisse einer Faktorenanalyse zeigten, dass die Jugendlichen einerseits in bestimmten Bereichen, wie zum Beispiel der makroskopischen Perspektive auf die Halbwertszeit, konsistent argumentierten. Andererseits zeigten die Ergebnisse einer Clusteranalyse, dass das konzeptuelle Verständnis der Jugendlichen in einem Bereich unabhängig von ihrem Verständnis in anderen Bereichen ist. Weiterhin traten Fehlvorstellungen weitestgehend unabhängig vom (Vor-)Wissen der Jugendlichen auf. Die Ergebnisse bestätigten zudem die Machbarkeit und den Nutzen, Schülervorstellungen zu Strahlung und Radioaktivität in einem Schülerlabor quantitativ zu messen.

Insgesamt zeigten die Studien im Rahmen des S'Cool LAB Projektes, dass das Thema Teilchenphysik eine vielversprechende Gelegenheit bietet, Jugendlichen moderne Physik näherzubringen und dabei ihr Interesse und ihre Neugier zu berücksichtigen, ohne ihr konzeptuelles Verständnis zu beeinträchtigen. Hier bestätigt die hohe pädagogische Effektivität dieses einzigartigen Beispiels eines Schülerlabors dessen erfolgreiche Entwicklung, rechtfertigt die erforderlichen Ressourcen und liefert äußerst vielversprechende Einblicke in die Wirkmechanismen außerschulischen Lernens.

Contents

Abstract i

Zusammenfassung iii

Contents v

1 Introduction 1

2 Theoretical and empirical background 5

2.1 Out-of-school learning opportunities 6

2.1.1 Definition and spectrum 6

2.1.2 Goals of out-of-school learning opportunities 7

2.1.3 Outcomes of out-of-school learning 7

2.1.4 A special case: out-of-school hands-on learning labs ('Schülerlabore') 8

2.1.5 Impact of out-of-school hands-on learning labs 10

2.1.6 Hands-on experiments as a special feature of out-of-school learning labs 13

2.1.7 The 'meet-a-scientist' effect as a special feature of out-of-school learning 14

2.2 Educational outcomes of out-of-school learning labs 16

2.2.1 Interest 16

2.2.2 Self-beliefs 18

2.2.3 Curiosity 22

2.2.4 Enjoyment, helpfulness, and easiness 25

2.2.5 Cognitive Activation 27

2.2.6 Cognitive outcomes 29

2.2.7 Concept map of affective and cognitive outcomes in S'Cool LAB 30

2.3 Predictors of educational outcomes in out-of-school learning labs 33

2.3.1 Contextual Model of Learning by Falk & Dierking 33

2.3.2 Cognitive load theory 34

2.3.3 Novelty theory 36

2.3.4 Novelty influence and experience factors in S'Cool LAB 40

2.4 Students' conceptions of particles in fields, radiation and radioactivity 43

2.4.1 Students' conceptions of electrically charged particles in magnetic fields 44

2.4.2 Students' conceptions of radiation and radioactivity 45

3 Research gaps and research questions 55

3.1 Limitations of previous studies 55

3.2 Motivation for this research project 57

3.3 Research questions 57

3.3.1 Affective & cognitive effects and outcomes 57

3.3.2 Additional questions as part of the formative assessment of S'Cool LAB 58
3.3.3 Students' conceptions of ionising radiation and radioactivity 58

4 Design of S'Cool LAB and its learning activities 59
4.1 History and goals of S'Cool LAB 59
4.2 S'Cool LAB's educational offers 61
4.3 Design principles for the development of hands-on activities 62
4.3.1 Predict-Observe-Explain (POE) 63
4.4 Description of the S'Cool LAB Day programme 66
4.5 S'Cool LAB workshops 68
4.5.1 Electron Tube - The Basics of Particle Acceleration 69
4.5.2 Cloud Chamber - Build and Observe a Particle Detector 71
4.5.3 X-rays - Medical Applications and Pixel Detectors 73

5 Data preparation and analysis methods 75
5.1 Standard methods and parameters 75
5.1.1 Item mean, standard deviation and standard error 75
5.1.2 Item easiness (or item difficulty) 76
5.1.3 Item discrimination 76
5.1.4 Cronbach's alpha & Kuder-Richardson index 76
5.2 Data screening and preparation 77
5.2.1 Missing data 77
5.2.2 Normality 78
5.2.3 Coding errors, outliers and contaminants 79
5.3 Discovering structures in S'Cool LAB's educational data 80
5.3.1 Factor analysis 80
5.3.2 k-means cluster analysis 83
5.4 Analysis of gains & differences 84
5.4.1 Comparing group means with t-tests 84
5.4.2 Effect size Cohen's d 84
5.5 Multiple linear regression analysis 86
5.5.1 Sequential linear regression 87
5.5.2 Assumptions of linear regression analysis 88
5.5.3 Model fit parameter 91
5.6 Multilevel linear modelling 91
5.6.1 Assumptions of multilevel linear modelling and evaluation parameters 93
5.6.2 Model fit criteria 95
5.7 Bonferroni correction and the inflation of alpha errors 96
5.8 Sampling and population 97
5.9 Software and software parameters 97

6 Pilot study – testing and developing questionnaires 99
6.1 Study framework and sample 99
6.2 Overview of items and scales forming the questionnaires 100
6.3 Item and instrument analysis 103
6.3.1 Dispositional and situational physics interest and self-beliefs 104
6.3.2 Curiosity trait and curiosity state towards particle physics 107
6.3.3 Enjoyment, helpfulness, and easiness 109
6.3.4 Novelty experience 111
6.3.5 Cognitive load 112
6.3.6 Cognitive activation 115
6.3.7 Tutor support 116
6.3.8 Additional items 117
6.4 Development of concept test items 118
6.4.1 Electron tube workshop: the Lorentz force 118
6.4.2 Cloud chamber workshop: cosmic particles 121
6.4.3 X-ray workshop: irradiation vs. contamination 124
6.4.4 Selection of concept test items for the main study 126
6.5 Pre-post effects of the S'Cool LAB hands-on session 128
6.6 Summary of the pilot study and conclusions for the main study 129

7 Main study – evaluating the impact of S'Cool LAB 131
7.1 Study framework and research design 131
7.2 Sample 133
7.3 Item and instrument analysis 135
7.3.1 Interest and self-beliefs 135
7.3.2 Curiosity trait and state 140
7.3.3 Test-retest reliability of students' motivational variables and curiosity 143
7.3.4 Enjoyment, Easiness, Helpfulness 144
7.3.5 Novelty 147
7.3.6 Cognitive load 151
7.3.7 Cognitive activation 156
7.3.8 Tutor support 156
7.3.9 Conceptual understanding 157
7.3.10 School grades mathematics, physics and English skills 160
7.3.11 Previous experiences with particle physics 161
7.3.12 Previous experiences with hands-on experiments 162
7.3.13 Previous experience with out-of-school learning opportunities 164
7.3.14 Summary of the items and scale analysis of the S'Cool LAB main study 166
7.4 Pre-post effects of the S'Cool LAB hands-on session 168
7.4.1 Interest in physics 169

7.4.2 Physics self-beliefs ... 169
7.4.3 Curiosity towards particle physics ... 170
7.4.4 Conceptual learning ... 170
7.4.5 Summary with respect to the research questions ... 187
7.5 Correlations among S'Cool LAB's educational outcomes ... 191
7.6 Additional questions as part of the formative assessment of S'Cool LAB ... 194
7.6.1 Perception of the hands-on session compared to the guided tour at CERN ... 195
7.6.2 Perception of the S'Cool LAB workshops ... 196
7.6.3 Important aspects of students' trip to CERN ... 198
7.7 S'Cool LAB's educational outcomes: sequential linear models (SLMs) ... 198
7.7.1 SLM: Situational interest ... 199
7.7.2 SLM: Curiosity state particle physics ... 205
7.7.3 SLM: Perceived enjoyment and helpfulness ... 209
7.7.4 SLM: Situational self-beliefs ... 214
7.7.5 SLM: Perceived easiness ... 218
7.7.6 SLM: Perceived cognitive activation ... 221
7.7.7 SLM: Conceptual understanding ... 225
7.7.8 Overview of eight sequential linear models on educational outcomes ... 228
7.8 S'Cool LAB's educational outcomes: multi-level models (MLMs) ... 230
7.8.1 Need for multi-level modelling ... 231
7.8.2 Model specification and additional considerations ... 234
7.8.3 MLM: Situational interest ... 239
7.8.4 MLM: Curiosity state particle physics ... 242
7.8.5 MLM: Perceived enjoyment ... 244
7.8.6 MLM: Perceived helpfulness ... 246
7.8.7 MLM: Situational self-beliefs ... 249
7.8.8 MLM: Perceived easiness ... 251
7.8.9 MLM: Perceived cognitive activation ... 253
7.8.10 MLM: Conceptual understanding ... 255
7.8.11 Comparison of multi-level models and single-level models ... 258
7.9 Summary: modelling S'Cool LAB's educational outcomes ... 259
7.9.1 Summary with respect to the research questions ... 260

8 Study on students' conceptions of radiation & radioactivity ... 271
8.1 Why Multiple-True-False items? ... 272
8.2 Item development ... 273
8.3 Sample and handling of missing data ... 274
8.4 Item analysis ... 275
8.4.1 Interviews and expert discussion ... 275
8.4.2 Item easiness ... 276

8.5 Classification of item categories 280
8.6 Exploratory factor analysis 282
8.6.1 Factor 1: Basic knowledge 284
8.6.2 Factor 2: Sub-microscopic basis of the radioactive decay (S-mi) 285
8.6.3 Factor 3: Half-life = half the source (S-ma) 286
8.6.4 Factor 4: Irradiation vs. contamination 287
8.6.5 Factor 5: Chemical properties of sources and invisible absorber effects 288
8.6.6 Factor 6: Tangible macroscopic radiation effects (A-ma) 289
8.6.7 Factor 7: Potpourri of sub-microscopic misconceptions 290
8.6.8 Items that were not interpreted within the seven-factor solution 290
8.6.9 Correlations between the extracted factors 291
8.6.10 Summary of factor analysis and scale properties 292
8.6.11 Types of radiation types: ionising versus non-ionising 293
8.7 k-means cluster analysis 294
8.8 Summary with respect to the research questions 296

9 Discussion and conclusions 299
9.1 Summary and implications 299
9.1.1 Educational outcomes 299
9.1.2 Novelty influence factors and novelty experience factors 304
9.1.3 Students' conceptions about ionising radiation and radioactivity 309
9.2 Strengths and limitations 311
9.3 Conclusions and outlook for future research 313

10 Appendix 317
10.1 Pre-visit questionnaire of the pilot study (paper version, grey scale) 317
10.2 Post-visit questionnaire of the pilot study (paper version, grey scale) 327
10.3 Pre-visit questionnaire of the main study (online version, grey scale) 343
10.4 Inter-visit questionnaire of the main study (paper version, grey scale) 358
10.5 Post-visit questionnaire of the main study (online version, grey scale) 360
10.6 Overview of sample of the pilot study 381
10.7 Main study: additional item and instrument properties 382
10.7.1 Clustering of students: Countries, teachers, groups 382
10.7.2 Gender (SEX) 383
10.7.3 Age versus grade (AGE) 383
10.7.4 When did students fill out the online questionnaires? 384
10.7.5 Dispositional and situational interest in physics (IE1, IE2, IE3) 385
10.7.6 Dispositional and situational physics self-beliefs (SC1, SC2, SC3) 386
10.7.7 Disp. and situat. experiments self-beliefs (SCE1, SCE3, SCE3) 387
10.7.8 Curiosity trait (CT) 388
10.7.9 Curiosity state particle physics (CS1, CS2, CS3) 389

10.7.10 Conceptual understanding (UN1 & UN3) ... 390
10.7.11 School grades (MA, PH, EN grade) ... 403
10.7.12 Experience with hands-on experiments (EH) ... 404
10.7.13 Experience with out-of-school learning opportunities (EO) ... 404
10.7.14 Experience with particle physics (EP) ... 405
10.7.15 Cognitive activation (CA) ... 405
10.7.16 Enjoyment (ENJ), Easiness (EAS) & Helpfulness (HEL) ... 406
10.7.17 Tinkering support, lack of problems & setting orientation ... 407
10.7.18 Cognitive load experiments (CLE) ... 408
10.7.19 Tutor support (TS) ... 408
10.8 Summary of extreme values and multivariate outliers ... 409
10.9 Sequential linear modelling: supporting material ... 411
10.9.2 SLM Situational interest ... 412
10.9.3 SLM Curiosity state post ... 414
10.9.4 SLMs Enjoyment and helpfulness: correlations matrix ... 416
10.9.5 SLM Situational self-beliefs: correlations matrix ... 420
10.9.6 SLM Easiness: correlations matrix ... 422
10.9.7 SLM Cognitive activation: correlations matrix ... 424
10.9.8 SLM Conceptual understanding post: correlations matrix ... 426
10.10 Final questionnaire (MTF items including item easiness) ... 428
10.11 Items that were excluded because of ambiguity ... 440
10.12 Scale properties of the extracted factors ... 441
10.12.1 Factor 1 (16 items) ... 441
10.12.2 Factor 2 (10 items) ... 441
10.12.3 Factor 3 (6 items) ... 442
10.12.4 Factor 4 (6 items) ... 442
10.12.5 Factor 5 (10 items) ... 442
10.12.6 Factor 6 (6 items) ... 443
10.12.7 Factor 7 (4 items) ... 443
10.12.8 Items that were not interpreted within the seven-factor solution ... 443
10.13 Overview of electronic appendix ... 444
11 References ... 445
12 Acknowledgments ... 461
13 Curriculum vitae ... 463
14 Publications in the framework of this PhD project ... 465
15 Lebenslauf ... 466
16 Publikationen im Rahmen dieser Doktorarbeit ... 468

1 Introduction

Most students agree that science and technology are important and they hold a mainly positive attitude towards STEM subjects (OECD, 2006; Sjøberg & Schreiner, 2010). However, in most of the OECD countries, science and technology student numbers declined over the past decades and women are still strongly underrepresented in STEM (OECD Global Science Forum, 2006). Here, it seems that science lessons in formal education fail to harness students' positive attitudes. Instead, several STEM school subjects, especially physics and chemistry, are among students' least favourite subjects (Zwick & Renn, 2000). In particular, students' declining enjoyment of and interest in school science, their low science-related self-beliefs, their stereotypic images of science and scientists, and thus, their low interest in science and technology careers are worrying (Osborne, Simon, & Collins, 2003).

However, today's students will need scientific literacy and scientific reasoning skills to face the challenges of the 21st century, for example, the revolution of job markets due to the growing important of machine learning and robotics. Here, science and engineering occupations are among the safest careers with a low probability of computerisation because they require higher-order complex thinking skills and creative intelligence (Frey & Osborne, 2017). Therefore, acquiring scientific reasoning skills is most likely key to success for current and future generations of students (Binkley et al., 2012).

Previous studies derive recommendations on how to change this worrying development. For example, Sjøberg and Schreiner (2010) suggest to make science learning more relevant to students and to take students' interests into account. Rocard et al. (2007) highlight the importance of fostering students' self-beliefs in science, in particular, for female students. Others recommend more hands-on activities and inquiry-based learning, but also more field trips and stronger links between formal and informal science education (Boiko, Nistor, Kudenko, & Gras-Velazquez, 2019; Rocard et al., 2007). Similarly, Lyons and Quinn (2010) suggest bringing students in contact with scientists, for example, in the framework of out-of-school learning opportunities.

Indeed, informal learning opportunities can influence students' career interests (Fadigan & Hammrich, 2004). Here, out-of-school learning laboratories hold a unique position within the informal learning sector because they provide hands-on activities, often guided by scientists, with the goal of fostering students' interest in science.

Furthermore, these labs provide authentic learning environments, which are often related to modern science topics, and research equipment that is usually not available in schools

(Hempelmann & Haupt, 2014). However, previous studies suspect that the high amount of novelty in these unique settings is not entirely beneficial, as high levels of novelty can hinder learning and development (Cors, Robin, Mueller, & Kunz, 2017; Guderian, 2007; Itzek-Greulich, 2015). In particular, student characteristics, such as their prior interest and experiences, as well as students' perceptions of characteristics of out-of-school learning labs, such as students' perception of the organisation, cognitive load or support by volunteering scientists, are expected to influence the educational effectiveness of hands-on sessions in out-of-school learnings settings (Cors, 2016; Glowinski & Bayrhuber, 2011; Molz, 2016).

In the framework of this PhD project, S'Cool LAB, a new hands-on out-of-school learning laboratory for high-school students was developed at CERN, Geneva, Switzerland. In doing so, particle physics learning activities were designed to help students understand the physics and technologies of the largest particle physics laboratory in the world. In addition to giving insights into the working methods, technologies and research at CERN, workshops in S'Cool LAB aim to foster students' interest in physics, their self-beliefs and their scientific curiosity. To make optimal use of this unique education opportunity at CERN the development of learning activities was based on documented students' conceptions and took recent research results about, for example, interest, curiosity, cognitive activation and cognitive load into account.

Specifically, three hands-on workshops were developed, focusing on the acceleration of electrons, the detection of cosmic particles, and X-ray imaging. Together, these three workshops formed a half-day hands-on session for high-school students aged 16 to 19. More than 500 students took part in the quantitative S'Cool LAB evaluation study, which analysed the educational effectiveness of the S'Cool LAB hands-on sessions. To understand the relevance of student and laboratory characteristics, different types of novelty factors were assessed in addition, and used to model S'Cool LAB's educational outcomes.

Furthermore, S'Cool LAB was successfully established as CERN physics education research facility. In the framework of a second study on students' conceptions, students taking part in hands-on workshops also contributed to expanding our current understanding of students' conceptions about ionising radiation and radioactivity.

Structure of the thesis

Chapter 2 introduces the theoretical and empirical background of out-of-school learning opportunities including their potential educational outcomes as well as novelty factors that can affect these outcomes. Furthermore, our current understanding of students' conceptions of electrically charged particles in magnetic fields, ionising radiation and radioactivity is summarised.

Chapter 3 highlights limitations of previous studies on hands-on out-of-school learning laboratories, describes the goals and objectives of the S'Cool LAB project in more detail and derives the respective research questions for the studies conducted within this PhD project.

Chapter 4 describes the development of the out-of-school learning laboratory S'Cool LAB and its learning activities. This includes the description of the half-day hands-on session that formed the basis of the S'Cool LAB evaluation study.

Chapter 5 provides an overview of the statistical methods used to prepare and analyse the data generated by students who took part in S'Cool LAB's hands-on workshops.

Chapter 6 presents the design and results of a pilot study in which questionnaire items and scales were developed and tested to evaluate their suitability for assessing both S'Cool LAB's educational outcomes and relevant novelty factors.

Chapter 7 describes the S'Cool LAB evaluation study, which is also referred to as 'main study' in the following. In particular, the effects of S'Cool LAB hands-on sessions on students are presented and the results of single-level and two-level linear regression models on S'Cool LAB's educational outcomes are described. Based on these regression models, important student and laboratory characteristics are highlighted and interpreted within the novelty factor framework.

Chapter 8 describes the students' conceptions study. In particular, the results of a factor analysis and a cluster analysis are presented and interpreted by taking documented students' conceptions into account.

Finally, **Chapter 9** summarises the results of the S'Cool LAB evaluation study and the students' conceptions study described before and discusses their implications. After reflecting on the limitations and strengths of the S'Cool LAB studies, an outlook on future research needs is presented.

2 Theoretical and empirical background

Learning and development are complex processes that can take place at any time and place, inside or outside school. Individual learners need to register and process new information, compare it with their existing knowledge structures, and integrate the new information into their long-term memory. Furthermore, learning is a social process in which emotions, motivation and cognition are equally important and highly intertwined components (Meyer & Turner, 2002). What is a complex process in formal education is even more complicated in out-of-school settings where even more internal and external factors might contribute to the educational outcomes. Specifically, out-of-school learning opportunities are less structured, less formalised, and instead characterised by a high level of novelty, free choice and interesting activities. Hence, while out-of-school learning opportunities can have advantages over formal education, their complex settings require special care when aiming at measuring educational outcomes. Furthermore, psychological constructs, like, for example, interest or higher-level cognitive learning goals, such as conceptual understanding require a theory-guided conceptualisation to allow using questionnaires for assessing their development. Indeed, Renninger (2007) acknowledges the difficulty of studying educational outcomes of informal science learning (ISL) and calls for future research to understand these learning settings better: *"Because ISL is a relatively new field, and ISL settings are so seemingly varied, no single inductive model (hypothesized set of indicators) exists that can be used to predict participants' changed enjoyment and learning in the ISL context. […] A next step for ISL research is likely to include establishing underlying design principles that can be used to inform the study of enjoyment and learning".*

Thus, the following sections first introduce out-of-school learning opportunities in general and their unique characteristics (section 2.1). In particular, out-of-school learning laboratories are presented as a special case of out-of-school learning and characterised by their unique features as well as their educational potential and effectiveness (subsections 2.1.4 and 0). Second, the expected affective and cognitive outcomes relevant for out-of-school learning labs are described and conceptualised based on previous research (section 2.2). Third, factors that influence the learning processes, especially in out-of-school learning opportunities, are derived from museum research, novelty theory and cognitive load theory (section 2.2.7). Last, the knowledge of students' conceptions is essential for designing and evaluating learning activities. Therefore, section 2.4 provides a broad overview of students' conceptions relevant for this project.

2.1 Out-of-school learning opportunities

2.1.1 Definition and spectrum

In OECD countries, lower secondary students spent on average 905 hours per year on compulsory instruction in formal classroom settings (OECD, 2014) accounting for only 15% of students waking time[1]. In addition to this compulsory formal education, students experience many different opportunities to learn and develop out of the school context. These out-of-school learning opportunities are sometimes further differentiated into informal and non-formal learning (European Centre for the Development of Vocational Training, 2009). Whereas the term informal learning refers in particular to unintentional learning without structure, such as learning the word 'big bang' while watching comedy on TV, learners engage intentionally in so-called non-formal learning activities, for example, when taking part in a science summer camp. Non-formal learning activities are typically structured according to their learning objectives, time or support. However, Anderson and Ellenbogen (2012) argue, that this classification depends on the epistemology of learning used and is somewhat arbitrary because even informal learning places such museum offer structured activities in some way. Furthermore, Dierking, Falk, Rennie, Anderson, and Ellenbogen (2003) recommend avoiding the term 'informal science learning' because of its limitations and its lack to capture real-world learning. Therefore, in the following, only the distinction between school and out-of-school learning opportunities (OSLEOs) is used.

Considering the amount of waking time learners spend outside compulsory instruction in school, and the amount of research on learning outside school, there seems to be a consensus that out-of-school learning opportunities are important. In particular, they play an essential role in STEM education (Dierking et al., 2003), because they can offer especially discipline related, authentic learning experiences, which school education cannot provide due to constraints on equipment, personnel, or time. There is a broad range of activities (such as summer camps, after-school programmes, science clubs) offered in a wide range of settings (such as universities, research institutes, museums, science centres, schools, maker spaces). They differ in time, frequency, duration, focus, activity types (often inquiry-based, learner-centred hands-on experiences), and achievement goal (from cognitive goals, such as awareness or knowledge gain to affective goals, such as interest) (Chi, Dorph, & Reisman, 2015).

[1] assuming 8 hours of sleep per night

2.1.2 Goals of out-of-school learning opportunities

There are different views on what students can or should learn in out-of-school learning opportunities, what is meant by learning, and how the success of a learning opportunity can be measured. In the following, learning is mostly used for cognitive changes, whereas development refers to changes in affective variables. According to DeWitt and Storksdieck (2008), cognitive, affective and social outcomes are considered important and valuable for out-of-school learning opportunities. However, they point out that short-term interventions taking place at out-of-school learning opportunities are more likely to impact affective outcomes, such as interest, attitudes, or curiosity than long-lasting cognitive changes. Furthermore, DeWitt and Storksdieck (2008) conclude that the benefit of out-of-school learning does not lie in teaching complex concepts or facts but in offering *"opportunities for exploration, discovery, first-hand and original experiences"*. Braund and Reiss (2006) further highlight the importance of five unique features of out-of-school settings for science learning: *"1. Improved development and integration of concepts. 2. Extended and authentic practical work. 3. Access to rare material and to "big" science. 4. Attitudes to school science: stimulating further learning. 5. Social outcomes: collaborative work and responsibility for learning."*

In summary, the goals of out-of-school learning and development include (Anderson & Ellenbogen, 2012; Chi et al., 2015):

- **Cognitive domains:** awareness, knowledge, understanding of STEM concepts, skills such as the use of research tools and techniques, scientific reasoning skills, knowledge about careers
- **Affective domains:** attitude, engagement, appreciation, interest, curiosity, excitement, motivation, moral, STEM identity formation, career aspirations, image of science and scientists

2.1.3 Outcomes of out-of-school learning

Out-of-school learning opportunities can have positive effects on learners, both on affective and cognitive constructs even if they involve only short-term interventions (Chi et al., 2015; National Research Council, 2009). For example, Fadigan and Hammrich (2004) report that study participants mention out-of-school learning influencing their educational and career decisions, suggesting that students with out-of-school science learning experience are more likely to pursue science in school. In particular, participants might remember a positive science learning experience, which can foster their science-related self-beliefs. Indeed, Renninger (2007) suggests that *"the 'fun' of the ISL* [informal science learning]

context may provide its participants with opportunities to feel successful with science and to begin developing feelings of self-efficacy and/or identity with science." Furthermore, in a large study by the European Schoolnet across five countries, 75% of the students report that a higher number of field trips would make their science learning more enjoyable, while 54% say that field trips never happen in their STEM lessons (Boiko et al., 2019).

However, the type, quantity and quality of educational outcomes of out-of-school learning opportunities vary. Renninger (2007) reports that participants in a citizen science project about birds *"did not show gains in either their attitudes towards or conceptual understanding of science; they did, however, demonstrate "increased knowledge of the biology and habitat needs of cavity-nesting birds".* Renninger (2007) also attributes this to the distinctive features of out-of-school learning opportunities: *"It is not surprising that studies of ISL may not typically demonstrate gains in knowledge or interest, given the short duration of an interaction with an exhibit, lack of scaffolding to stretch thinking, and the emphasis on fun."*

Previous studies have also shown that the added value of out-of-school learning settings can be limited when comparing cognitive and affective outcomes of school and out-of-school learning settings by controlling for very similar hands-on learning activities (Itzek-Greulich et al., 2015, 2017). Hence, merely changing the learning setting is not beneficial and does not make use of the unique potential of out-of-school learning. Another factor that can limit cognitive learning in these settings is an overwhelming amount of novelty and choice. For example, Bamberger and Tal (2007) compare free, limited and no choice activities in a museum and their effects on students' learning. They report that the limited choice setting provides scaffolding leading to a higher level of control and deeper engagement and in turn supports learning.

Because the nature of learning in out-of-school contexts is very complex (Anderson & Ellenbogen, 2012), Dierking et al. (2003) recommend researchers studying outcomes of these learning settings to take several aspects into account such as the specific physical setting, sociocultural aspects (e.g., by taking group effects into account in multilevel models), the learning process itself in addition to the outcomes, and the variety of learning and development outcomes (e.g., by including them in hierarchical linear models).

2.1.4 A special case: out-of-school hands-on learning labs ('Schülerlabore')

A special type of out-of-school learning opportunities for high-school students are hands-on learning labs (in German: 'Schülerlabore' (Hempelmann & Haupt, 2014)). Over the past 20 years, hundreds of these labs have been founded in Germany, Austria and Switzerland

by science research institutes, research centres and universities (Hempelmann & Haupt, 2014). As an interactive way of science communication, outreach and education, they specialise on offering hands-on activities using high-tech equipment in an authentic environment to introduce high-school students to modern scientific research.

In many cases, the topics of out-of-school learning labs are challenging to teach in schools because of their complexity and abstractness, but also because of the lack of affordable hands-on experiments and demonstrations. By funding research equipment for students through outreach budgets, out-of-school learning labs enable students to work independently on different experiments guided by volunteering scientists. Often, this hands-on experimentation is combined with onsite visits to the 'real' research facilities. One of the common goals is to foster students' interest in science, technology, engineering, or mathematics. However, specific goals range across science communication, teacher training, science education research, and attracting future trainees and employees (Haupt et al., 2013; Hempelmann & Haupt, 2014; Streller, 2015). Table 1 summarises the similarities between out-of-school learning labs, formal, and informal science learning.

Table 1. Characteristics of different learning settings based on European Centre for the Development of Vocational Training (2009), Hempelmann & Haupt (2014), Wellington (1990) and Streller (2015).

	Formal science learning e.g., traditional school lesson	**Science learning in out-of-school hands-on learning labs**	**Informal science learning** e.g., family visit to the museum
Focus of activities	teacher-led, teacher-centred	learner centred	learner-led, learner-centred
Instructor	teacher	tutors, often scientists	none, or guide
Target audience	students in classes	students in classes	mostly individuals
Structure	highly structured objectives and time	structured objectives and time	unstructured
Openness	more closed	guided	open-ended
Topic	curriculum	connection to curriculum	no restriction
Assessment and certification	yes	no	no
Social aspect	less central	central	central
Main activities	listen, read, write, calculate	Independent hands-on experimentation	read, observe, hands-on exhibits
Participation	mandatory	in part voluntary	voluntary
Learning material	blackboard, books, teacher	experiments and instructions	labels, texts, objects

Since many of these hands-on labs welcome whole classes of students together with their teachers and provide structured activities for small groups of students, they have much in

common with formal education activities. However, the absence of grading pressure, social aspects, the presence new topics, setting and material and a higher degree of free choice correspond more to informal learning as, for example, in the museum context. Therefore, it is useful to combine research results from both formal and informal learning settings when informing the design and evaluation of learning and development in out-of-school learning labs.

2.1.5 Impact of out-of-school hands-on learning labs

Do out-of-school hands-on learning labs have an impact on their participants? And, if so, which effects were measured by previous research studies? Table 2 provides an overview of studies on out-of-school science learning labs conducted in Germany.

Table 2. Results of previous studies on out-of-school science learning labs in Germany based on (Mokhonko, Nickolaus, & Windaus, 2014; Molz, 2016; Streller, 2015). Positive effects (+), negative effects (-), no effects (0), and studied influence factors (x) are indicated for students' situational interest and self-beliefs as well as their achievement.

Authors	Subjects: Physics (P), Chemistry (C), Biology (B)	School grades	Situational interest	Situational self-beliefs	Knowledge gain	Understanding	Novelty factors
(Engeln, 2004)	P, C	9-10	+				x
(Scharfenberg, 2005)	B	12	0		+		
(Brandt, 2005)	C	7-8	+	+			
(Guderian, 2007)	P	5-8	+				
(Glowinski, 2007)	B	12	+				
(Zehren, 2009)	C	8-10	+	+	+		
(Pawek, 2009)	P	9-13	+	+			x
(Weßnigk, 2012)	P, C	10-11	+	+			
(Itzek-Greulich, 2015)	C	9	+		+		
(Mokhonko, 2016)	P, C	7-10	+	+			
(Huwer, 2015)	C	5,10	+	+	+		
(Streller, 2015)	P	10-13	+				x
(Molz, 2016)	P	9-13	+	+	+	+	x
(Buse, 2018)	B	11-13		+	+		x
(Rodenhauser, 2018)	B	9-13	+	+			
(Simon, 2019)	P	9-13	+	+		+	x

In summary, previous studies confirm short-time effects on affective and cognitive educational outcomes. Almost all of the labs can **trigger students' situational interest**.

Furthermore, there seems to be no gender difference in this effect (Engeln, 2004; Guderian, 2007). In a study with more than 11 000 students, Priemer, Menzl, Hagos, Musold, and Schulz (2018) report an average value of 59% for the epistemic component of situational interest and confirm a negligible gender difference (Cohen's $d_C = 0.1$).

However, this situational interest typically fades within a few months and rarely stabilises into longer-lasting dispositional interest. Mokhonko et al. (2014) report only very small effects on female learners' dispositional interest and dispositional self-beliefs in physics (Cohen's $d_C = 0.1 - 0.2$). To stabilise these effects, studies investigated multiple visits (Guderian, 2007; Zehren, 2009) or embedding the out-of-school learning experience into school learning settings. Moreover**,** adding preparation and post-visit activities leads to heterogeneous and small effects (Glowinski, 2007; Guderian, 2007; Streller, 2015; Zehren, 2009). For example, Streller (2015) confirms positive effects on students' situational interest and self-beliefs when preparing them via an online portal. He highly recommends mandatory preparation and post-visit activities to enhance or maintain the effects of a one-time visit to an out-of-school learning lab. However, Itzek-Greulich et al. (2015) report that the learning outcome is independent of integration into classroom.

Several authors suspect that the high amount of **setting novelty** in out-of-school learning labs affects learning and development (Cors et al., 2017; Guderian, 2007; Itzek-Greulich, 2015). However, only few studies take students' perceptions of their novelty experience into account or consider factors that might influence novelty experience. For example, Pawek (2009) identified support by tutors and the positive learning atmosphere as promising predictors for situational interest, whereas Engeln (2004) lists cognitive challenge and comprehensibility. Simon (2019) investigated the impact of the quality of support and reports that the perceived cognitive activation, clear instructions and supportive tutors are important predictors of students' outcomes. Molz (2016) studied novelty effects on cognitive and affective outcomes systematically. He confirmed that preparation aiming at reducing setting novelty has positive effects on curiosity and motivation. However, when comparing treatment groups that differed only by learning place and explainer, both Itzek-Greulich et al. (2017) and Molz (2016) report no effect on affective outcomes.

Although out-of-school learning labs typically invest a lot of time and effort into designing their hands-on workshops, most studies on out-of-school learning labs focus only on affective outcomes (Table 2). A handful of studies confirm that these experiences produce **cognitive effects** as well, in particular, when measuring knowledge gain. For example, Buse (2018) reports an impressive effect on German students' knowledge gain in

neurobiology independent of whether the one-day workshop was offered bilingual or in German (Cohen's $d_C = 2.0$). She even confirmed a high level of interconnectedness of the newly acquired knowledge by analysing concept maps. Also, Huwer (2015) reports small to medium-sized effects of participation in hands-on workshops on students' chemistry knowledge (Cohen's $d_C = 0.3 - 0.5$ for different topics). Moreover, Simon (2019) reports a reduced prevalence of certain misconceptions after participation in hands-on workshops.

Can out-of-school learning experiences produce better cognitive results than formal instruction in schools? Itzek-Greulich (2015) analysed the differences between similar one-day learning activities taking place in school, in an out-of-school lab, and in an embedded experience with school and out-of-school component. Compared to the control group, all treatment groups performed better on a paper and pencil knowledge test. However, the group that experienced only learning activities in school outperformed the other groups, which supports the claim that out-of-school learning opportunities are not ideal when aiming for cognitive effects. The author attributes her findings to a high amount of novelty hindering learning in the novel out-of-school setting through cognitive overload. Molz (2016) reports no difference between the cognitive outcomes of hands-on activities taking place inside or outside school. However, pre-and post-visit lessons had a significant effect on students' achievement. Indeed, students taking part in embedded hands-on activities at school outperformed students taking part in isolated hands-on activities outside school. These findings suggest, that visiting out-of-school learning labs only makes sense, if they offer activities that cannot be done at school, for example, by providing authentic research equipment guided by 'real' scientists.

In summary, out-of-school learning labs can have cognitive effects together with affective effects, although it seems that the regular school setting better supports learning. However, most of the employed achievement tests focus solely on learning of facts in biology and chemistry contexts that happens as a 'side-effect'. Molz (2016) is the first to include items from validated concept tests sensitive to testing students' conceptions and understanding of physics concepts. Since physics, in particular, suffers from low interest and self-beliefs, and is perceived as particularly difficult, more research and repetition of previous research is needed for higher-order learning in physics out-of-school learning labs.

Overall, out-of-school learning labs might be relatively expensive outreach tools, but they have an immense potential to influence students through unique learning and development experiences. However, their impact seems to depend on the quality of the intervention, as well as on several novelty factors, and the mixed results of prior research require further investigation. Moreover, all above-mentioned studies took place in Germany with German

students, and it is not clear to what degree their findings can be generalised. More research about educational outcomes and the mechanism of action of novelty factors will greatly improve our understanding of successful out-of-school science learning labs.

2.1.6 Hands-on experiments as a special feature of out-of-school learning labs

What distinguished hands-on learning labs from other out-of-school science learning opportunities is their focus on hands-on experimentation. Why experiments? Already Pestalozzi (1826) promoted true, natural and meaningful education through a holistic learning approach involving the heart, mind and hand ('Herz, Geist und Hand'). Since then, only very few science educators questioned the necessity of experiments. Nonetheless, the discussion of the role and educational effects of lab work in science education has a long tradition, and learning from hands-on experiments is (still) one of the central problems of physics education research (Abrahams & Millar, 2008; Dillon, 2008; Hodson, 1988; Hofstein & Lunetta, 2004; Trumper, 2003).

According to Hofstein and Lunetta (2004), the **goals of hands-on laboratory activities** include:

- Learning about scientific concepts/mastering subject matter
- Learning about the nature of science
- Learning about scientific reasoning
- Development and improvement of practical & problem-solving skills
- Fostering interest and motivation

However, there are several **problems with hands-on activities** that might hinder reaching these goals. When students follow detailed cookbook-like instruction, they rarely activate deeper level processing and instead process new information only at a surface-level (Baird 1990). In this way, conceptual learning is impossible. However, students often have too little prior knowledge to handle more open and complex experimental situations, which would provide an even higher cognitive load (Hart, Mulhall, Berry, Loughran, & Gunstone, 2000). Therefore, the best many students can do is to follow provided instructions (Hodson, 1993). Furthermore, learners often perceive following instructions as the purpose of their lab work, together with 'getting the right answer' they fail to link theory and experiment (Hofstein & Lunetta, 2004). Hofstein and Kind (2012) conclude that "*[t]he biggest challenge for practical work, historically and today, is to change the practice of 'manipulating equipment not ideas'. The typical laboratory experience in school science in a hands-on but not a minds-on activity.*"

Even though there is a long tradition of empirical research, there is **no clear evidence for the benefit of hands-on activities** with regard to cognitive and affective outcomes. When comparing cognitive and affective effects of hands-on experiments with demonstrations or lessons without experiments, previous studies show mixed results (Garrett & Roberts, 1982; Harlen, 1999; Hodson, 1990, 1996; Hofstein & Lunetta, 2004; Hopf, 2007; Woolnough & Allsop, 1985). For example, Woolnough and Allsop (1985) report that girls show very little interest in hands-on activities compared to boys, whereas other studies report that these activities can even reduce the gender gap in interest (Burkam, Lee, & Smerdon, 1997; Pawek, 2009). Although Hodson (1990) sees little educational value in the type of practical work conducted in many schools, students seem to prefer practical work over other activities in school (Edwards & Power, 1990; Swarat, Ortony, & Revelle, 2012). Students even wish for more hands-on experiments to make their science lessons more enjoyable (Boiko et al., 2019).

Indeed, the question remains, whether effects of hands-on work can be measured at all (Hofstein, Shore, & Kipnis, 2004), especially considering the wide range of different hands-on activities (Atkinson, 1990). Nevertheless, research shows that well designed hands-on activities have great potential in science education (Hofstein & Lunetta, 2004).

Here, **suggestions to improve hands-on activities** and to increase their educational effectiveness include:

- Concentrate on fewer and more important tasks (Hodson, 1993)
- More conceptual questions and encouraging communication between learners (Cox & Junkin III, 2002)
- Authentic contexts (Molyneux-Hodgson, Sutherland, & Butterfield, 1999)
- The right amount of cognitive load (openness versus instructional guidance) (Carbonneau, Marley, & Selig, 2013; Kirschner, Sweller, & Clark, 2006)
- Minds-on as well as hands-on: cognitive activation/challenge through connecting and reflecting on theory and practical work (Abrahams & Millar, 2008; Gunstone & Champagne, 1990; Hodson, 1993)
- Enable students to reflect on their prior conceptions (Gunstone & Champagne, 1990; Miller, Lasry, Chu, & Mazur, 2013)

2.1.7 The 'meet-a-scientist' effect as a special feature of out-of-school learning

Another special characteristic of hands-on learning labs is the contact to scientists. Many facilities rely on volunteers from their scientific community to facilitate their hands-on activities (Hempelmann & Haupt, 2014). This contact to 'real' scientists instead of teachers

can have positive effects on students' attitudes towards science, because it might change their perception of scientists (Houseal, Abd-El-Khalick, & Destefano, 2014; Rennie & Howitt, 2015; Woods-Townsend et al., 2016). Instead of being confronted with stereotypes, students meet 'real' and normal people that are typically much younger and potentially more accessible role models than, for example, Marie Curie and Albert Einstein. Indeed, except for their teachers, students rarely meet scientists in formal education. However, **talking to scientists in out-of-school settings can greatly influence students' educational and career decisions** (Fadigan & Hammrich, 2004). Here, already an extremely short invention of five minutes, which included information about activities of scientists in their everyday work life, provided first evidence of an increase in students science career aspirations (Hochberg & Kuhn, 2019).

Research about teachers confirms enthusiastic teaching characterised by teachers' enjoyment and excitement as a very effective instructional feature that leads to positive effects on students' cognitive and affective outcomes (Keller, Neumann, & Fischer, 2013; Kunter et al., 2008). In particular, students' affective outcomes, such as enjoyment correlate with students' perception of teacher enthusiasm (Frenzel, Goetz, Lüdtke, Pekrun, & Sutton, 2009). However, mediated by students' attention and other affective variables, this might also influence cognitive outcomes (Keller et al., 2013). Scientists are typically very enthusiastic about their work and their research topic, especially when talking to students or the public. For instance, in a study by Stylinski, Storksdieck, Canzoneri, Klein, and Johnson (2018), 285 scientists taking part in an online public engagement training evaluated the priority of their public engagement objectives. The goal with the highest agreement was indeed *"getting people excited about science"*. In summary, connecting students and scientists is a way to enhance students' perception of relevance and excitement in science learning (Tytler, Symington, & Clark, 2017).

Hence, even though scientists' pedagogical knowledge and their pedagogical content knowledge can be assumed to be lower than that of experienced science teachers, their enthusiasm is a promising factor when evaluating the potential of out-of-school learning labs. Pawek (2009) confirms that the perception of tutoring scientists in out-of-school learning labs is a strong predictor of the triggered situational interest. Also, Glowinski and Bayrhuber (2011) identified the support by scientists ('quality of instruction') as important predictor of situational interest in experiments. However, the path coefficient of the direct effect in their path model for the situational interest in experiments was only significant for students with a low dispositional interest, whereas students with a high science interest seemed to appreciate the support by scientists less.

2.2 Educational outcomes of out-of-school learning labs

There are several possible affective and cognitive outcomes of out-of-school learning (subsection 2.1.3), and many of them are interlinked. Hence, Anderson and Ellenbogen (2012) as well as Gutman and Schoon (2013) recommend studying the variety of learning and development outcomes instead of isolated concepts. In the following, several educational outcomes that are relevant for the S'Cool LAB project are described, starting with a definition of the respective construct, its relevance for out-of-school learning opportunities and links to other constructs as well as its conceptualisation and quantitative measurement in previous studies. Finally, a concept map of all described concepts and their interconnectedness is presented (subsection 2.2.7).

2.2.1 Interest

Definitions of interest

Interest defines a psychological state that occurs and develops during interactions between persons and their environment (e.g., objects, topics, contents, ideas). This **situational interest** is characterised by focused attention and an affective reaction, such as enjoyment (Hidi & Baird, 1986; Hidi & Renninger, 2006; Krapp, 2002; Schiefele, 1978). The term interest also refers to a relatively long-lived predisposition to engage with certain ideas or learning content, which is often referred to as individual or **dispositional interest**. Dispositional interest develops over time and differs between people.

Development of interest: catch and hold facets

Krapp (2002) describes three phases of interest development: triggered situational interest, stabilised situational interest, and dispositional (or individual) interest.[2] Although situational interest and dispositional interest are different constructs, it is assumed that a high situational interest triggered by specific characteristics of a learning environment (so-called 'catch-facets') can transform into more stable interest constructs under the right conditions (so-called 'hold-facets') (Krapp, 2002).

According to Mitchell (1993), **catch-facets** include any instructional tool that gets the interests of the students, such as interesting equipment (e.g., computers), interesting activity types (e.g., group work) or interesting tasks (e.g., puzzles). According to Rotgans

[2] Similar to this 3-phase model of interest development, Hidi and Renninger (2006) propose a 4-phase model that *"describes 4 phases in the development and deepening of learner interest: triggered situational interest, maintained situational interest, emerging (less-developed) individual interest, and well-developed individual interest"*.

and Schmidt (2014), a perceived knowledge deficit can also create situational interest, which in turns means that situational interest decreases with increased knowledge about the topic. Furthermore, surprising information or intensity are listed as triggers of situational interest by Hidi and Renninger (2006).

Mitchell (1993) further describes how the perceived meaningfulness of the learning activities for students' daily lives can 'hold' up students' interest and potentially help to internalise interest (**hold-facets**). Besides, his research results suggest that involvement in the learning process and students perceiving themselves as active participants in the learning process, for example, by doing hands-on activities on their own instead of listening to a lecture, is a vital tool of enhancing and maintaining situational interest. Schraw and Lehman (2001) support his results by suggesting that *"active involvement is an important variable in catching and holding interest. Presumably, anything that is done to promote active involvement (e.g. promoting deeper processing) should increase interest."*

Interest and out-of-school learning labs

Interest in a school subject such as physics can be based on the learning content itself or on the specific learning setting, which includes interesting activities, such as experimenting or debating (Krapp, 2002). Both stable and malleable components of interest are interesting for out-of-school learning activities: Indeed, stable interest in a topic correlates positively with the conceptual understanding of physics texts independent of students' prior knowledge (Andre & Windschitl, 2003) and with conceptual change (Mason, Gava, & Boldrin, 2008). However, even students that show the more short-lived situational interest demonstrate a higher persistence and deeper level processing of information, which is in turn associated with improved comprehension and learning (Ainley, Hidi, & Berndorff, 2002; Ainley, Hillman, & Hidi, 2002; Schraw & Lehman, 2001). Indeed, Rotgans and Schmidt (2018) report only situational interest as a significant predictor of knowledge acquisition whereas dispositional interest merely influences the situational interest learners display when starting a task.

In general, interest and academic achievement typically show weak to moderate positive relationships (Schiefele, Krapp, & Winteler, 1992). The PISA 2006 results, for example, show a weak correlation of $r = 0.13$ between science interest and science performance for the pooled OECD sample (OECD, 2007). Interest is also an important aspect of course and career choices (Hidi, 2006; Krapp, 2002; Schraw & Lehman, 2001).

Furthermore, previous research indicates that students' learning motivation, interest and attitudes towards school, especially towards school physics, decrease over time (Todt &

Schreiber, 1998; Wigfield & Eccles, 1992). Moreover, this effect seems to be stronger for girls than for boys (Hoffmann, Häußler, & Peters-Haft, 1997). Therefore, out-of-school learning opportunities and in particular, out-of-school learning labs should use their potential to foster students' interest by providing activities triggering a high situational interest, ideally independent of students' previous knowledge, dispositional interests and gender. Ultimately, students might actualise their already existing dispositional interests or develop their situational interest into dispositional interest when they engage longer in a topic and perceive it as meaningful (Krapp, 2002).

Conceptualisation and measurement of interest

Interest involves both affective components and cognitive components, which continuously interact with each other. Whereas affects are often associated with the situation that triggers interest, the cognitive component of interest becomes more important when interest develops into a disposition over time (Hidi, 2006). In particular, interest is associated with positive feelings, increased perceived value and increased knowledge or desire for knowledge about the object of interest (Hidi, 2006; Krapp, 2002; Renninger, 2007). The epistemic component of interest drives a person to find out more about a certain object (and is closely related to curiosity, see subsection 2.2.3). The value-related component refers to the personal significance of the object of interest, whereas the feeling-related component refers to a positive experiential state with feelings, such as enjoyment, involvement, stimulation, fascination, or sometimes flow (Krapp, 2002).

Questionnaires trying to assess students' interest in a particular school subject (e.g., physics) or activity (e.g., hands-on workshops) are therefore typically composed of items about the epistemic component (e.g., *"I want to learn more about …"*), value-oriented items (e.g., *"I find it important to learn about …"*), and feeling-oriented items (e.g., *"I like my physics lessons"*) (Fechner, 2009; Kuhn, 2010; Molz, 2016).

2.2.2 Self-beliefs

Definition of self-beliefs

There are different theoretical approaches to academic self-perceptions or self-beliefs. The **self-concept of ability** is a relatively stable mental representation of oneself, which includes a collection of cognitive beliefs about oneself formed through experience and feedback from the environment (Gutman & Schoon, 2013; Valentine, DuBois, & Cooper, 2004; Wild & Möller, 2009). Affective self-beliefs (e.g., *"I like physics"*) are typically associated more with interest than with self-concept (Wild & Möller, 2009). The **global self-concept of one's ability**, sometimes called self-esteem, includes perceptions and

feelings about oneself in general, while the **domain-specific self-concept** (e.g., academic self-concept) refers to perceptions of oneself in a specific area (e.g., in school) (Wild & Möller, 2009).

Marsh (1990) provides evidence for a hierarchical multi-faceted academic self-concept, which separates math and verbal academic self-concepts and lists, for example, the subject-related self-concepts for physics and mathematics as direct components of the math academic self-concept. Factor analysis confirmed that students report separate self-concepts for their school subjects. However, a second-order factor analysis supported the assumption that the academic self-concept can be explained to a certain degree by the two main factors verbal and mathematics self-concept. Therefore, Marsh recommended using subject-specific scales when measuring a specific component of academic self-concept.

While self-concept is formed through past experiences, **self-efficacy beliefs** are cognitive competence expectations oriented at future context-specific tasks. Because of their task dependence, self-efficacy beliefs are not as stable as self-concept (Bandura, 1977; Wentzel & Miele, 2016). However, despite the difference in their theoretical conceptualisation, it is methodologically challenging to measure self-concept and self-efficacy beliefs separately (Valentine et al., 2004). Therefore, they are often described together forming the **subject related self-beliefs**.

Self-beliefs and out-of-school learning labs

According to Schunk and DiBenedetto (2016), there are several open questions that limit our understanding of the development of self-beliefs in out-of-school learning settings: *"How do students learn and refine their skills and develop self-efficacy outside of formal instructional contexts? What roles do other persons in these environments, such as parents, peers, mentors, and coaches, play in helping students become self-efficacious learners?"*

Research within formal education contexts demonstrate reciprocal effects between the academic self-concept and academic achievement and recommends fostering both simultaneous (Marsh & O'Mara, 2008; Valentine et al., 2004); higher self-beliefs predict higher levels of academic achievement and vice versa. For example, Hattie (2008) reports an effect size of self-concept on achievement of $d_C = 0.42$. Also, academic self-beliefs (together with interest and achievement) are important predictors of students' high-school course choices and career aspirations (Nagy et al., 2008; Parker, Nagy, Trautwein, & Lüdtke, 2014)

Furthermore, Skaalvik and Skaalvik (2004) report that academic self-beliefs differ with gender and age with respect to mathematics: *„Male students had higher self-concept, performance expectations, intrinsic motivation, and self-enhancing ego orientation in mathematics than did female students, whereas female students had higher intrinsic motivation for learning language than did male students. Older students had higher verbal than mathematics self-concept regardless of gender."* In a study with students from Australia, the US and Germany, Nagy et al. (2010) confirm the same gender differences starting at grade 7. They further report a decline of maths self-beliefs with age. Pajares and Miller (1994) even used the concept of self-beliefs to mediate the influence of gender and prior experience on mathematics performance. Women and men also seem to differ in the way they create STEM self-beliefs with men using mastery experience as source whereas for women *"social persuasion and vicarious experiences were the primary source"* (Zeldin, Britner, & Pajares, 2008). Overall, the differences – if they exist – are small to moderate: *"Despite performing equally well as males in most countries, females tend to have a weaker self-concept in science than males. However, this difference remains moderate, ranging from a strong effect in some countries to no difference in others"* (OECD, 2007).

In summary, out-of-school learning opportunities should use their unique role to foster not only students' achievement or interest through suitable learning activities but also their academic self-beliefs through challenging but achievable tasks promoting students' feeling of competence. Previous meta-analyses confirm the effectiveness of targeted interventions. For example, Haney and Durlak (1998) report a medium-sized average effect size from pre-test to post-test of $d_C = 0.57$ for programmes aiming to improve the self-concept. A more recent meta-analysis by O'Mara, Marsh, Craven, and Debus (2006) reports a similar mean effect size of $d_C = 0.51$ for the impact of self-concept interventions, with a higher effect size of $d_C = 0.67$ for interventions targeting self-concept development directly (versus indirectly $d_C = 0.48$).

Interest, self-beliefs and the self-determination theory of motivation

Motivation in the education context is *"typically defined as a set of interrelated desires, goals, needs, values, and emotions that explain the initiation, direction, intensity, persistence, and quality of behaviour"* (Wentzel & Miele, 2016). Wild and Möller (2009) describe motivation as a psychological force, which is central to understanding students' learning success. Deci and Ryan's empirical self-determination theory (SDT) assumes that *"humans have a natural tendency to learn and develop, as reflected in the dual tendencies toward intrinsic motivation, the natural tendency and energy to engage in behaviors that*

people find challenging and interesting, and integration, the tendency to internalize and assimilate social practices and mores, even those they may not find interesting" (Ryan & Deci, 2016). SDT was built on research on the differences of intrinsically and extrinsically motivated actions and the social conditions, which affect intrinsic motivation and integration. Extrinsic factors, such as rewards, but also externally imposed controls, such as deadlines, surveillance of behaviour, or competition undermine intrinsic motivation because they negatively affect the sense of self-determination. Hence, extrinsically motivated learning is no longer caused by curiosity and interest, and it tends to be less thorough (Deci & Ryan, 1981).

Their theory describes different levels of motivation based on the perceived degree of self-determinedness of an individual. Nonself-determined behaviour is associated with amotivation, higher levels of self-determinedness are associated with extrinsic motivation, and self-determined behaviour is associated with intrinsic motivation. Furthermore, they list interest, enjoyment and inherent satisfaction as relevant regulatory processes for intrinsically motivated behaviour (Ryan & Deci, 2000).

SDT also describes the conditions (or psychological needs) that affect the development of motivation as well as the strength of motivation: the need to feel competent, autonomous, and related to others. Here, fostering autonomy in the learning process does not mean *"abandoning all structures, but rather creating structures that allow freedom within limit"* (Deci & Ryan, 1981). For example, a meta-analysis reports a positive effect of choice on intrinsic motivation (Patall, Cooper, & Robinson, 2008).

If these basic needs are satisfied, individuals act interest-oriented and autonomously motivated, which is associated with high task performance (Deci & Ryan, 2002; Deci & Ryan, 2012). Interest and self-beliefs are often considered facets of learning motivation. Hence, it is no surprise that the psychological basic needs of self-determination theory partially overlap with the hold-facets of interest theory. Furthermore, the basic need for feeling competent relates to the construct of self-beliefs. Also, interest characterises intrinsic motivation. Wild and Möller (2009) even describe interest as a source for intrinsic motivation because in an interest-based situation a person is doing what they want to do. Therefore, all interest-based actions are per definition self-determined (Krapp, 2002). However, these two constructs are not identical. For example, interest does not necessarily include decision-making or a drive to act and it can change in interaction with the environment (Renninger, 2007).

Conceptualisation and measurement of academic self-beliefs

Several survey instruments have been used to measure self-concept. Marsh (1990) developed the Academic Self-Description Questionnaire (ASDQ), which includes six items for every school subject with a six-level rating scale. The items are based on prior experiences of performance in the subject (e.g., *"I have always done well in [a specific school subject]."*), comparison with other (e.g., *"Compared to others my age I am good at [a specific school subject]"*) and perceived competence feeling (e.g., *"Work in [a specific school subject] classes is easy for me"*).

2.2.3 Curiosity

Definition

Defining the construct curiosity is challenging, as there exist many different and partially inconsistent conceptualisations. Curiosity is often used synonymously with interest or other similar constructs, such as the need for cognition, intellectual engagement, openness to experience, sensation seeking, or wonder (Grossnickle, 2016). Following a review of 39 articles, Grossnickle (2016) identifies **four themes for definitions of curiosity**:

- Curiosity as a need for knowledge, information, or learning something new; learners need to be aware of existing knowledge gaps
- Exploratory behaviour as defining outcome of curiosity
- Curiosity triggered by collative variables (e.g., novelty, complexity, ambiguity, challenge, disequilibrium, uncertainty) especially in case of unexpected or counterintuitive phenomena
- Characterized by (positive or negative) emotions or arousal

Subsequently, Grossnickle (2016) defines curiosity *"as the desire for knowledge or information in response to experiencing or seeking out collative variables, which is accompanied by positive emotions, increased arousal, or exploratory behavior"*. Furthermore, Grossnickle (2016) identifies different types of curiosity based on the object of curiosity, its degree of stability, and its reason.

Object of curiosity: Berlyne (1978) introduced two types of curiosity, both associated with increased arousal. While **perceptual curiosity** describes the drive of an individual to experience and feel triggered by sensory stimulation, the **epistemic curiosity** describes the drive to know, to seek opportunities to engage intellectually, to acquire knowledge and facts about a certain topic. In addition, Grossnickle (2016) describes **social curiosity** as desire to know more about other people and **physical curiosity** as desire to explore and manipulate oneself and one's surroundings.

Stability: Whereas **trait curiosity** is defined as the tendency of an individual to seek for new knowledge and to be dispositionally curious about a range of topic, **state curiosity** can be seen as resulting curiosity triggered by collative variables attracting attention to a certain topic (Schmitt & Lahroodi, 2008). These collative variables are properties of the stimulus determining the level of cognitive arousal, such as novelty, surprisingness, incongruity, complexity, uncertainty, conflict (Berlyne, 1960, 1978). Although trait and state curiosity are separate constructs, they are related with correlations between $r = 0.24$ (Litman, Hutchins, & Russon, 2005) and $r = 0.78$ (Reio Jr, Petrosko, Wiswell, & Thongsukmag, 2006).

Reason: **Interest-type curiosity** is associated with a desire to learn new things accompanied by positive emotions, such as enjoyment. **Deprivation-type curiosity** involves negative emotions regarding the identified knowledge gaps and results in a need for information to close these gaps, which then leads to positive emotions (Litman & Spielberger, 2003).

The difference between epistemic curiosity and interest

The psychological construct of epistemic curiosity is very similar to other constructs, such as 'need for cognition' (Cacioppo & Petty, 1982), or the constructs interest and engagement. It is challenging to study curiosity without these related constructs (Arnone, Small, Chauncey, & McKenna, 2011; Berlyne, 1978; Deci & Ryan, 1981; Schmitt & Lahroodi, 2008), it is even debated whether curiosity and interest can be used synonymously (Kashdan & Silvia, 2009).

Situational interest and specific epistemic curiosity can both be considered motivational constructs that encourage to interact with the environment with the goal to obtain more information. Thereby both are relevant when trying to improve motivation to learn (Alexander & Grossnickle, 2016). Indeed, both constructs seem to be natural tendencies of humans, and both concepts are sensitive to environmental factors, such as novelty. However, the constructs of interest and curiosity differ in a number of ways, namely in their **direction of focus and in the different types and stability** (Grossnickle, 2016).

Direction of focus: Interest is generated in the interaction between a person and the environment; therefore, interest is always identified with a specific topic (e.g., physics), it is not a global construct. However, Naylor (1981) describes a global construct of curiosity trait which is independent of a topic. Instead, curiosity is directed at perceived knowledge gaps which can occur in a variety of topics. Grossnickle (2016) further highlights the value of curiosity compared to interest: *"By virtue of its interest independence, curiosity leads us*

to topics outside our specialty and even overrides the practical and epistemic interests and the motivationally original desires that limit our inquiry to our specialty. This keeps us in touch with the wider world and occasionally brings to our attention topics outside our specialty that bear on inquiry within our specialty."

Although both curiosity and situational interest can be caused by a state of conflict and uncertainty, some topics (such as death or violence) seem to trigger emotional interest universally without needing an additional trigger.

Types, stability and development: Whereas both constructs have trait as well as state components, so far only the development of interest from situational interest towards a more stable predisposition has been described (Hidi & Renninger, 2006; Krapp, 2002). There is no theory yet to describe the sequential development from the short-lived curiosity state into trait, and researchers are missing an intermediate and more enduring form of curiosity (Grossnickle, 2016). Although trait curiosity affects the experience of state curiosity (Litman et al., 2005), trait curiosity is seen as very stable over time and across different contexts, whereas interest is always directed towards a specific object/topic and therefore more likely to change over time.

In summary, Alexander and Grossnickle (2016) recommend measuring both interest and curiosity as *"the understanding of either of these motivational constructs cannot be examined in isolation from the other"*. However, it is not clear if it is always possible and necessary to separate the constructs when using questionnaires.

Curiosity and out-of-school learning labs

Von Stumm, Hell, and Chamorro-Premuzic (2011) identify epistemic trait curiosity as one of the direct predictors of academic achievement (with a path weight of 0.2) together with general intelligence and conscientiousness. Curiosity enhances memory of surprising new information (Kang et al., 2009) and is correlated with higher academic performance when using standardised tests (Wavo, 2004). Indeed, when curiosity is high, the pupils of study participants were dilated (Eckstein, Guerra-Carrillo, Singley, & Bunge, 2017). This marks the release of the neurotransmitter noradrenaline in the brain, which facilitates learning (Aston-Jones & Cohen, 2005).

Although most curiosity research took place in formal learning settings, Arnone et al. (2011) point out, that studying curiosity only in formal learning settings might limit our understanding of this construct since learning motivation (including curiosity) declines over school time (Harter, 1981). For this reasons, out-of-school learning opportunities may be crucial to maintain or even foster curiosity in science and can act as fruitful research

settings: *"Research focused on out-of- school learning contexts may provide a clearer picture of the role of curiosity and the maintenance of interest and engagement in learning and offer important directions for how to address the lack of student motivation often present in today's classrooms."* (Arnone et al., 2011).

Furthermore, Schmitt and Lahroodi (2008) highlight the potential of hands-on inquiry-based learning to foster curiosity early on: *"Schooling is sometimes blamed for the evaporation of curiosity, but the mounting practical concerns that accompany maturation may take the larger toll on curiosity, discouraging firsthand inquiry not instrumental to those concerns. Unfortunately, curiosity is just what is needed to form a practice of firsthand inquiry that can prevent its degeneration. For this reason, the best opportunity for mature curiosity lies in the early formation of a practice of firsthand inquiry, relying on immature curiosity."* In addition to fostering curiosity by offering suitable learning activities, the characteristics of the learning environment can increase students' epistemic state curiosity. Studies demonstrated, for example, the impact of text structure and type of writing task (Boscolo, Ariasi, Del Favero, & Ballarin, 2011; Knobloch, Patzig, Mende, & Hastall, 2004) or the positive effect of controversy with peers (Lowry & Johnson, 1981).

Conceptualisation and measurement of curiosity

Curiosity is mostly measured through self-reports, and several different scales exist focusing on different aspects of curiosity, see Grossnickle (2016) for an overview. For example, the Melbourne Curiosity Inventory (MCI) focuses on general trait and state curiosity in educational settings (Naylor, 1981). Litman and Spielberger (2003) studied meaningful dimensions of the multifaceted personality construct of curiosity. They identified two correlated factors corresponding to concepts suggested by Berlyne (1978): the concept of epistemic curiosity (EC, the 'drive to know') and perceptual curiosity (PC, the 'drive to experience and feel'). Litman and Spielberger (2003) further describe their epistemic curiosity scale and its factor structure, which consists of diversive components (seeking a broad range of information) and specific components (interest in learning detailed knowledge about a specific topic). Hirth (2019) adapted existing curiosity trait and state scales to assess the effects of smartphone-supported learning activities about acoustics phenomena on German high-school students.

2.2.4 Enjoyment, helpfulness, and easiness

A common goal of out-of-school learning opportunities is enjoyment, experiencing pleasure while exploring without feeling overwhelmed (Allen, 2004; Martin, 2004). Rennie (1994) measured affective outcomes from visits to hands-on learning experiences at science education centres. She argues that both affective and cognitive outcomes of out-of-school

learning places are important: an enjoyable visit might lead to further cognitive learning, whereas maximum cognitive learning requires a positive affective experience. In addition to enjoyment, Rennie (1994) describes helpfulness and easiness as components of students' affective outcomes.

Enjoyment, helpfulness, easiness and other affective variables

Enjoyment and interest are sometimes used interchangeably (Renninger, 2007), and indeed, enjoyment is often considered a component or a requisite of interest. For example, the PISA 2006 study assessed the 'enjoyment of science' as one of the measures of students' intrinsic motivation to learn science and report very strong correlation between the 'enjoyment of science' scale and the 'general interest in science' scale ($r = 0.88$) (OECD, 2007). The results of the PISA 2006 study also confirm that, in general, students enjoy learning science independent of their gender (with the exception of small gender differences in a few countries, e.g., Japan, the Netherland and Korea). Furthermore, enjoyment of learning science was positively associated with performance in most countries ($r = 0.19$ for the pooled OECD sample) (OECD, 2007).

Helpfulness measures the perceived usefulness of the activities for the students. Hence, it is closely related to value-oriented component of interest and serves as an indicator for a hold-component of interest development. Here, correlations to the situational interest are expected, as well as correlations with enjoyment: If students enjoy learning science, they are also more likely to perceive it as meaningful (Gläser-Zikuda & Mayring, 2003).

The perceived **easiness** is related to the cognitive load of the activities but also to students' self-beliefs. Too much enjoyment and perceived easiness could, however, limit the effects of out-of-school learning opportunities, because students might not perceive these activities as 'real science' (Archer et al., 2010; Carlone, 2004). Therefore, perceived easiness should not be too high to make sure activities cognitively activate students through challenging but doable tasks.

Conceptualisation and measurement of enjoyment, helpfulness and easiness

Rennie (1994) presents a questionnaire with the subscales enjoyment, usefulness and perceived self-competence, which are three separate but related components as confirmed by factor analysis. She also reports a small correlation ($r = 0.32$**) between cognitive and the affective enjoyment scale.

2.2.5 Cognitive Activation

Definition

From a constructivist viewpoint, conceptual understanding can only be reached through an active and elaborated interaction with the learning topic (Lipowsky, 2009). Sometimes, authors refer to 'active' interactions in case of purely behaviourally active activities, such as hands-on experiments (Chi & Wylie, 2014). However, cognitively activating tasks refer to tasks that require deeper level processing of information and can, of course, include suitable hands-on activities.

Craik and Lockhart (1972) describe **different levels of processing** from a memory psychology perspective with their levels-of-processing framework: the deeper a certain learning content is processed, the better it will be remembered. With the word 'depth' the authors refer to a higher degree of semantic or cognitive analysis of a stimulus. Deep levels are thus associated with pattern recognition and meaning-making, whereas low levels include, for example, repetition. Therefore, learners comparing and connecting new information to their prior knowledge will be more successful than learners who repeatedly receive the same visual information. Indeed, Craik and Lockhart (1972) considered using the word 'spread' instead of 'depth', because, according to their interpretation, deeper processing levels also refer to a higher level of connectedness of new information.

The construct of cognitive activation is a comparatively new but crucial dimension for assessing instructional quality (Klieme, Pauli, & Reusser, 2009; Lipowsky, 2009). However, it is used almost interchangeably with the above described 'deep level processing'. Effective teaching strategies to promote deeper level processing include providing cognitive challenges in the classroom, for example, by contrasting students' existing conceptions with new surprising experiences, which allow students to connect their knowledge in new ways or apply it in new contexts. Furthermore, critical thinking and problem-solving tasks require deeper level processing, as well as tasks aiming at structuring information. Other strategies are targeting the learning atmosphere and materials. Students, who can discuss their ideas about an interesting topic in their classroom, will be more engaged and more active in their learning process (Kunter et al., 2013).

Cognitive activation and out-of-school learning labs

Cognitive activation is a predictor of learning gains (Lipowsky et al., 2009). In mathematics education, it is associated with higher achievement while also supporting students' characteristics, such as motivation, self-efficacy and self-concept (Burge, Lenkeit, &

Sizmur, 2015). As a feature of instructional quality, cognitive activation mediates the effects of teachers‘ pedagogical content knowledge on students’ learning gains (Baumert et al., 2010; Kunter et al., 2013). However, the effect of deeper level activation depends on the learning outcome tested. In a quasi-experimental study, Hänze and Berger (2007) compared the jigsaw classroom method of cooperative instruction with traditional direct instruction. They measured differences in students' experience of the three basic needs of self-determination theory (autonomy, competence, and social relatedness), self-reported cognitive activation, and students’ degree of intrinsic motivation. They suspect that measuring factual knowledge gains, which are associated with shallow cognitive processing, are not suitable to show the benefits of cognitive activation. Instead, they suggest testing conceptual understanding rather than recalling of facts to capture effects of deeper level learning processes.

Cognitive activation, interest and performance

Although interest and cognitive activation are related, it is not clear if cognitive activation leads to situational interest and intrinsic motivation, or whether it is a characteristic of learning activities dominated by interest. For example, Lipowsky (2009) describes cognitive activation as mediator between interest and learning outcomes. However, Förtsch, Werner, Dorfner, Kotzebue, and Neuhaus (2017) describe the classroom criterion cognitive activation as predictor of situational interest (standardised $\beta = 0.2$**). Hence, students’ self-reported cognitive activation can also be considered a separate educational outcome of learning activities, but higher levels of situational interest should be correlated with higher levels of cognitive activation.

In any way, both interest and cognitive activation are interesting predictors of learning. Intrinsic motivation is associated with the use of more effective deep level processing strategies and deeper levels of comprehension (Schiefele, 1991). Indeed, Benware and Deci (1984) compared conceptual and rote learning in an active learning and a passive learning setting: ‘Active’ learners were asked to learn a topic and to teach it later to other students, whereas ‘passive’ learner were asked to learn a topic to be tested later. Although both groups of students were only tested on their learning outcomes, students in the ‘active’ setting showed higher conceptual learning, were more intrinsically motivated, and perceived themselves as more actively engaged. However, the groups showed no difference in rote learning. Similarly, Schiefele (1991) reports significant correlation between performance and interest only for performance test items requiring a deeper level of processing, whereas a negative correlation was found between interest and rehearsal. Moreover, Hofer, Schumacher, Rubin, and Stern (2018) find that students in cognitively

activating physics lessons outperform students in regular lessons both in conceptual understanding and quantitative problem-solving performance.

Conceptualisation and measurement of cognitive activation

Cognitive activation cannot be observed directly. Furthermore, it is used both as a feature of instructional quality on the classroom level and as a student-level variable. Studies by Benware and Deci (1984) or Hofer et al. (2018) compare the educational outcomes of learning settings designed to foster cognitive activation and normal learning settings. Hence, they use cognitive activation as a feature of instructional quality but do not measure at the student level if this assumption is correct. Baumert et al. (2010) investigate cognitive activation as a feature of instructional quality by analysing tests, examinations and homework assignments of mathematics teachers. Klieme et al. (2009) analyse video recordings of mathematics lessons.

Jurik, Gröschner, and Seidel (2014) studied students' self-reported cognitive activation using ten Likert-type items, e.g., *"During the last two lessons, I explained the content in my own words"*. Hänze and Berger (2007) developed a reliable scale with five items (Cronbach's $\alpha_C = 0.75$) to assess 'activation of deeper level processing' based on Wild (2000) that includes items about students' cognitive engagement (*"I felt focused."*, *"I felt involved in learning the material."*) as well as items derived from suitable deep-level learning strategies, such as critical review (*"I took a critical look at the new material."*), structuring information *("I tried to distinguish between important and unimportant things."*) and elaboration (*"I tried to connect what I was learning with things I already knew.").*

2.2.6 Cognitive outcomes

Definition

Although many of the affective outcomes described above have a cognitive component (for example, self-beliefs), cognitive outcomes typically refer to knowledge gains or a better conceptual understanding of the learning topic.

Cognitive outcomes of out-of-school learning labs

Most out-of-school learning opportunities *"do not have explicit aims for participant learning. Rather, they are designed to satisfy participants' aims to enjoy cultural, social, and/or educational experiences"* (Renninger, 2007). Instead, out-of-school science activities might produce cognitive outcomes as side effects of students' engagement (Falk & Dierking, 2016). However, cognitive outcomes can be achieved, see also subsection 2.1.3. If the conditions are favourable, museum visits can even foster conceptual learning (Martin, 2004). In general, it seems that out-of-school settings have been underestimated

as learning opportunities. Indeed, previous research on the cognitive outcomes of out-of-school learning labs is limited and heterogeneous, see subsection 2.1.5. Typically, the out-of-school learning settings do not outperform formal learning in the school settings if the learning content can be taught in the same way in school. Specifically, Molz (2016) finds that a well prepared and reinforced day with hands-on lab activities at school can lead to higher cognitive outcomes than a poorly prepared visit to an out-of-school hands-on lab. In summary, the power of out-of-school learning labs seems to lie in unique learning activities supported by special equipment that cannot be offered in formal learning settings.

Conceptualisation and measurement of cognitive outcomes

Cognitive outcomes are usually measured through kind of paper-and-pencil tests before and after the activities. While factual knowledge can be measured easily by evaluating students' answers to simple recall tasks, it is difficult to assess students' level of conceptual understanding. Usually, sophisticated and meticulously designed concept tests such as the FCI (Hestenes, Wells, & Swackhamer, 1992) or the BEMA (Chabay & Sherwood, 1997a) are employed to measure students' conceptual learning.

2.2.7 Concept map of affective and cognitive outcomes in S'Cool LAB

The previous subsections highlight different educational outcomes and their relevance for out-of-school learning labs. In particular, the outcomes described above were assessed within the S'Cool LAB project. But why is it worthwhile to study several outcomes in parallel and how are these outcomes connected?

Anderson and Ellenbogen (2012) describe different epistemological and historical perspectives on out-of-school learning research: While early studies tried to reduce the complexity of learning to a few well-defined and easily measurable variables by removing contextual factors, later studies took the context of these learning settings into account. The 'Contextual Model of Learning' (Falk & Dierking, 2000) is one example for the latter type of research (see subsection 2.3.1). Although the broader view and deep level research on individuals and their experience (which is influenced by different contexts) led to a much better understanding of a variety of desirable outcomes, this type of research was limited to case studies (Anderson & Ellenbogen, 2012). Since small-scale qualitative studies focusing on particular programme outcomes cannot offer systematic or generalizable evidence, studies that take into account the complex nature through suitable analysis methods are needed. Here, Anderson and Ellenbogen (2012) emphasise the need to *"build and test common theoretical frameworks"*. In this spirit, this section connects different educational outcomes within a concept map based on the literature review in

section 2.2 to highlight their intertwined nature and the difficulty of establishing causal relationships between the different variables.

Physics interest and self-beliefs are central constructs for learning processes as they are both facets of learning motivation. While interest is a mostly affective construct, self-beliefs have stronger cognitive components. However, both constructs are usually significantly correlated (e.g., $r = 0.56$ in a study by Marsh, Trautwein, Lüdtke, Köller, and Baumert (2005)).

Interest is significantly correlated with enjoyment and activation (Schiefele, 1991). Indeed enjoyment can be seen as a component of interest. However, the direction of the relationship between cognitive activation and interest is not quite clear. Cognitive activation is often described as an aspect of interest (sometimes a hold-facet (Mitchell, 1993)), and previous studies have shown that interested-oriented learning corresponds to higher levels of cognitive activation (Krapp, Renninger, & Hidi, 1992). Dispositional interest influences situational interest (together with the catch-facts of a learning activity), and situational interest can develop into a more stable dispositional interest over time. The latter process is supported by hold-facets, such as meaningfulness (similar to helpfulness) or cognitive activation. Furthermore, the epistemic component of interest is almost indistinguishable from epistemic curiosity. Interest also correlates with different cognitive learning outcomes (Schiefele et al., 1992), such as conceptual understanding.

Self-beliefs develop with age and are based on previous experiences. Hence, the easiness of a learning activity contributes to updating the self-beliefs. Self-beliefs correlate with measures of achievement such as conceptual understanding through reciprocal effects (Marsh & O'Mara, 2008). Even if male and female student do not differ in physics school grades or achievement, female students usually report significantly lower physics self-beliefs (Skaalvik & Skaalvik, 2004).

Figure 1 summarises the connections between these educational outcomes in a concept map. The map also includes dispositions and more stable constructs that affect the development of their situational counterparts.

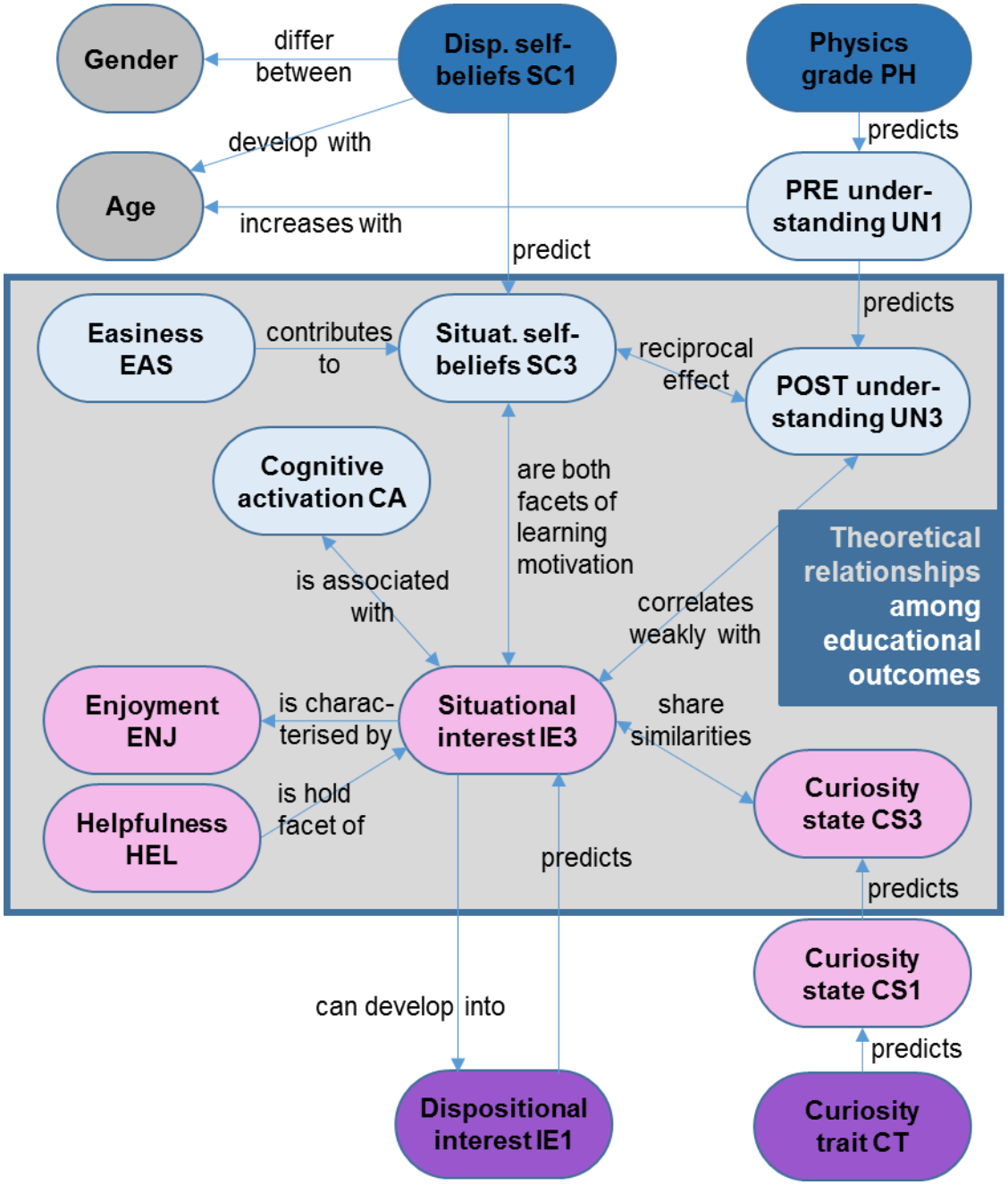


Figure 1. Concept map of educational outcomes. In blue: rather cognitive constructs, in purple: rather affective constructs. Darker shades indicate more stable constructs/dispositions/traits.

2.3 Predictors of educational outcomes in out-of-school learning labs

Learning and development are complex processes in school and out-of-school. Numerous different factors influence what and how students learn, and how their affect changes. When focusing on out-of-school settings such as museums or school trips, which are typically characterised by an immense amount of additional novelty (such as unfamiliar places and equipment, free choice, explainers), one could argue that these settings are even harder to understand than school settings (Osborne & Dillon, 2007). Consequently, research projects focusing on out-of-school settings are confronted with additional challenges. Whereas classrooms allow randomised control group designs to manipulate and study variables in quasi-experimental methods, challenges of research on informal learning settings range from problems with data taking to issues with comparing research results from these highly specialised but uncontrollable settings. However, there exist attempts to structure the influence factors. This section introduces different views and aspects of the rich variety of out-of-school learning settings starting with the Contextual Model of Learning (Falk & Dierking, 2000; Falk & Dierking, 2016) as a framework to structure influence factors (subsection 2.3.1), followed by an overview of cognitive load theory (2.3.2), and novelty theory as well as its application and measurement in out-of-school learning settings (2.3.3). Finally, subsection 2.3.4 summarises the novelty factors relevant for the S'Cool LAB project.

2.3.1 Contextual Model of Learning by Falk & Dierking

Falk and Dierking (2000) propose the Contextual Model of Learning as a framework for cognitive learning and affective development in the museum context. They distinguish three main contexts that affect learners' outcomes:

- The **personal context** refers to what learners bring with them to the informal learning place as individuals. It includes their prior knowledge and experiences, their interests, their motivation and expectations.
- The **sociocultural context** refers to social and cultural factors affecting learning. The cultural background of the learners can, for example, affect the value attributed to out-of-school learning. Learning further depends on social interaction within a group of learners (peers, family) or with outsiders (explainers, other visitors).
- The **physical context** refers to the physical characteristics of the informal learning setting: the architecture of the building, the exhibits and their features in a museum, the equipment learners can interact with, the learning materials and activities offered to the learners.

- In addition, **time** is described as relevant for all learning and development processes and reflects also the influence of preparation and follow up activities.

Falk and Dierking (2000) describe their model as a framework: *"a device for organizing the complexities of learning within free-choice settings"*. However, they do not suggest which factors are most important in a given situation or how to measure all these factors. Consequently, Falk and Storksdieck (2005) derive 12 key factors from these contexts and previous research results that have been shown to influence the quality of museum learning (see Table 3). For example, learners' prior knowledge, experiences and interest influence learning in museum as well as learners' orientation in the museum or the design of activities.

Table 3. 12 key factors influencing the quality of museum learning by Falk and Storksdieck (2005).

Personal context *1. Visit motivation and expectations* *2. Prior knowledge* *3. Prior experiences* *4. Prior interests* *5. Choice and control* *Sociocultural context* *6. Within group social mediation* *7. Mediation by others outside the immediate social group*	*Physical context* *8. Advance organizers* *9. Orientation to the physical space* *10. Architecture and large-scale environment* *11. Design and exposure to exhibits and programs* *12. Subsequent reinforcing events and experiences outside the museum*

Besides, Falk and Storksdieck (2005) study the correlation between 24 measures, representing 11 of the 12 listed key factors, and changes in seven different learning measures. Although their data supports the Contextual Model of Learning, the effect of the independent variables was easiest to interpret when analysing more homogenous sub-samples based on prior interest and knowledge. It seems that certain characteristics of the learners influence how they are affected by the characteristics of the out-of-school learning environment. Furthermore, these factors have not been derived from an underlying theory but have been selected based on empiric evidence. Here, cognitive load theory and novelty space theory can provide this missing theoretical basis and are therefore introduced in the following.

2.3.2 Cognitive load theory

Based on the model of working memory (Baddeley & Hitch, 1974) and in particular its limited capacity (Miller, 1956), Sweller (1994) developed the cognitive load theory for learning processes, which takes different factors of task difficulty into account. He describes the level of cognitive load through the level of 'element interactivity'. While

‘elements’ consist of any concept or procedure that needs to be learned or has been learned previously, ‘element interactivity’ refers to the connections between these elements. Learning separate elements with only a few connections to other elements (such as names of elementary particles) is considered a ‘low element interactivity’ task. In contrast, ‘high element interactivity’ material consists of numerous connected elements, such as hands-on physics experiments requiring learners to activate and connect knowledge about different concepts while manipulating novel equipment to observe and interpret experimental outcomes.

Types of cognitive load

The cognitive-load-theory distinguishes three types of cognitive load: **Intrinsic cognitive load** refers to the intrinsic complexity of the presented information. Educators cannot control this component of cognitive load as it originates from the complex learning subject itself. If it is not possible to further reduce the element interactivity of the learning material, the remaining cognitive load is intrinsic to the learning content. However, the level of intrinsic load depends on the learners and their level of expertise, because chunking of information into higher-order schemata will minimise the cognitive effort required to process this information.

Extraneous cognitive load originates from the presentation of learning content, which does not contribute to learning. Inappropriate instructional procedures may impose an additional extraneous cognitive load. It is desirable to reduce the element interactivity of the learning material as much as possible without changing the learning content (e.g., by removing purely decorative pictures or additional ‘decorative’ descriptions), to occupy as little of the learners’ mental capacity as possible with extraneous cognitive load.

Germane cognitive load describes the proportion of working memory resources that is devoted to the essential learning process, such as the construction and automation of schemata by interpreting, classifying, differentiating and organising. Effective learning tasks increase the proportion of germane cognitive load by stimulating students to engage in such learning processes.

Although the construct of cognitive load plays a central role in education research, the clarity of its components has been criticised previously (De Jong, 2010; Schnotz & Kürschner, 2007). For example, high levels of perceived cognitive load can be interpreted as beneficial or detrimental for learning – depending on the proportion of germane versus extraneous cognitive load. However, it is not possible to distinguish these two constructs empirically. Also, De Jong (2010) argues that the amount of cognitive load required to

construct a schema is independent of the accuracy of this schema: Learners might perceive a high level of cognitive load while reinforcing existing alternative conceptions.

Conceptualisation and measurement of cognitive load

Apart from physiological measures such as heart rate variation or functional MRI scan, self-ratings through questionnaires can be used to assess cognitive load (Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

The NASA Task Load Index (NASA-TLX) is a multidimensional scale that has been developed by Hart and Staveland (1988) to assess the perceived subjective workload. It produces a weighted average of the six subscales mental demands, physical demands, temporal demands, performance, effort and frustration. Cors (2016), Hochberg (2016) and Molz (2016) adapted this scale to fit the context of out-of-school learning labs and to measure an (unweighted) overall score for the perceived cognitive load. Items refer, for example, to mental demand (*"How high was the level of mental stress you experienced"* [very low/very high], effort (*"How hard did you have to work to perform the experiments?"* [very little/very much]), frustration (*"How frustrated were you?"* [very little/very much]), and temporal demand (*"How do you rate the duration of the different experiments?"* [very short/very long]). However, due to its multidimensionality, the scale's internal consistency is typically low (from $\alpha_C = 0.59$ (Hochberg, 2016), to $\alpha_C = 0.63$ (Molz, 2016) and $\alpha_C = 0.70$ (Cors, 2016)). In the context of an out-of-school learning lab, Hirth (2019) uses an additional set of items to assess the cognitive load associated with using experimental equipment (in particular with using smartphones for measurements).

2.3.3 Novelty theory

Novelty describes the properties of a stimulus causing arousal, curiosity, and subsequently explorative behaviour (Berlyne, 1978). As such, novelty is crucial to all learning and development activities, and Falk, Martin, and Balling (1978) summarise: *"Novelty, and the very powerful needs for exploration it generates, is an extremely important educational variable".* Out-of-school learning opportunities are typically characterised by a very high amount of novelty. However, too much novelty can hinder learning by cognitively overloading learners' working memory capacity or by promoting negative emotions, such as fear.

The novelty space model by Orion & Hofstein (1994)

Falk et al. (1978) first suggested that setting novelty influences students' behaviour and cognition in out-of-school learning settings. Later, Orion and Hofstein (1991) studied

different factors influencing the educational outcomes of geologic field trips. These factors included student-level variables, teaching variables, and field trip variables (Table 4).

Table 4. Factors influencing the educational outcomes of field trips (Orion & Hofstein, 1994).

Student factors	Teaching factors	Field trip factors
Previous knowledge	Curriculum structure	Learning conditions
Acquaintance with field trip area	Didactic methods	Duration
Previous experience with field trips	Learning material	Attractiveness
Previous attitudes	Quality of teachers	Weather conditions
Gender		
Class characteristics (grade, size, major studies)		

Subsequently, Orion and Hofstein (1994) suggest that two main factors influence the educational effectiveness of field trips: the quality of the learning opportunity (for example, the structure and quality of learning materials) and the 'novelty space', which is defined by at least three 'novelty factors' that represent different student factors as listed in Table 4:

- **Cognitive**: previous knowledge about the learning content
- **Psychological**: previous experience with field trips as learning opportunities
- **Geographic**: familiarity and orientation with the field trip area

Their research results further show that students receiving preparation to reduce their novelty space outperform students who receive no preparation when measuring students' attitudes and achievement after the field trip (Orion & Hofstein, 1994).

Out-of-school learning model by Anderson and Lucas (1997)

Anderson and Lucas (1997) introduce a variation of this model for out-of-school learning linking perceived novelty, curiosity and learning. They suggest that, instead of a minimum level of perceived novelty, an optimal level of perceived novelty (not too low, not too high) will lead to the highest learning outcomes (Figure 2). However, Anderson and Lucas (1997) consider only 'cognitive learning', whereas Orion and Hofstein (1994) consider cognitive outcomes (achievement) as well as affective outcomes (attitude).

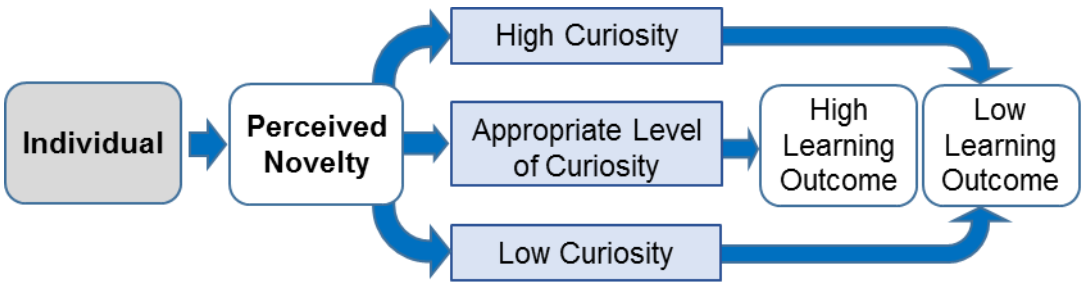


Figure 2. The learning model as a function of novelty/curiosity (Anderson & Lucas, 1997).

Novelty factor framework by Cors et al. (2017)

Based on a literature review of eight studies about novelty at out-of-school learning opportunities, Cors et al. (2017) summarise that *"optimized novelty relates to positive educational outcomes"* while *"too much or too little novelty can be problematic*". Furthermore, they build onto the Contextual Model of Learning and the novelty space model to present a comprehensive framework for studying novelty and its impact on affective and cognitive educational outcomes in these learning settings. In particular, Cors et al. (2017) differentiate 'novelty influence factors' (NIFs) and 'novelty experience factors' (NEFs).

Novelty influence factors are student characteristics that affect how students perceive novelty at an out-of-school learning opportunity, such as dispositions, previous knowledge, and experiences. In this way, Cors et al. (2017) extend Orion and Hofstein's novelty space by deriving five types of novelty influence factors from empirical evidence (see Figure 3):

- **Cognitive**: previous knowledge about the learning content, school grades
- **Affective**: attitudes towards the out-of-school learning experience, and impressions of the learning experience
- **Setting familiarity**: familiarity and orientation with the out-of-school learning opportunity
- **Social**: relationships with peers or educators
- **Capabilities**: skills needed for the out-of-school learning activities

Novelty experience factors, on the other hand, are factors indicating students' perceived novelty during an out-of-school learning opportunity, such as exploration behaviour, state of curiosity, setting orientation, or cognitive load. Novelty experience factors can be influenced by students' novelty influence factors (as suggested by Falk and Storksdieck (2005)) and might in turn influence the educational outcomes. For example, if students receive sufficient preparation about the learning content in advance, their perceived cognitive load at the out-of-school learning opportunity will be lower, which might consequently foster learning.

In summary, the novelty factor framework by Cors et al. (2017) helps to structure the numerous underlying theories (e.g., person-object-theory of interest, self-determination theory, cognitive load theory, novelty space model) and to interpret the impact of out-of-school learning for both affective and cognitive educational outcomes by identifying important novelty factors predicting these effects. While novelty influence factors are essentially student factors, novelty experience factors are indicators of students'

perception of the quality of the out-of-school learning opportunity (see teaching and field trip factors suggested by Orion and Hofstein (1994), see Table 4). The latter can also be used to identify potential areas of improvement of out-of-school learning opportunities. In principle, the five types of novelty influence factors could also translate into the same five types of novelty experience factors, see Figure 4, although some novelty experience factors may be linear combinations of several factors.

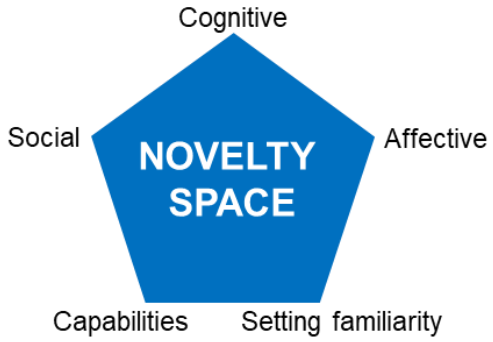


Figure 3. Extended novelty space with five types of novelty influence factors (Cors et al., 2017).

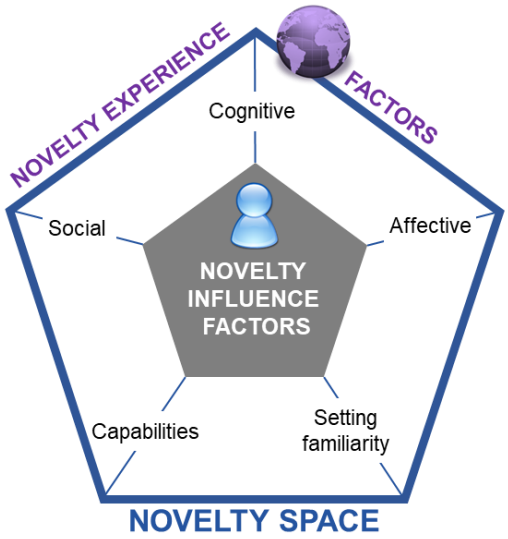


Figure 4. Extended novelty space with five types of novelty influence and experience factors.

Impact of novelty on educational outcomes of out-of-school learning laboratories

As discussed in subsection 2.1.5, previously, only a few studies take students' perceptions of their novelty experience into account or consider factors that might influence students' novelty experience. For example, Pawek (2009) identifies the social factors 'support by tutors' and a positive 'learning atmosphere' as promising predictors for situational interest, whereas Engeln (2004) emphasises the importance of the cognitive factors 'cognitive challenge' and 'comprehensibility'. Simon (2019) reproduces these results by highlighting the importance of clear instructions, support by tutors, a positive learning atmosphere and students' perceived cognitive activation.

Molz (2016) studied novelty effects on cognitive and affective outcomes systematically and confirms that preparation and post-visit activities lead to higher achievement. Furthermore, students participating in a well-embedded laboratory session at their school outperform students participating in a very similar out-of-school laboratory session without preparation and post-visit activities. Also, preparation aiming at reducing setting novelty has a positive effect on curiosity and motivation contrary to the belief that too much information before the visit would decrease students' curiosity.

Cors (2016) studied preparation, novelty influence factors, novelty experience factors and their impact on students' educational outcomes of the so called mobiLLab programme[3], which comprises a one-day hands-on session with a mobile high-tech laboratory visiting schools. She finds that the number of novelty-reducing videos students watched for preparation did not predict students' novelty experience. However, the three assessed novelty influence factors (technological capabilities, experience with out-of-school learning, and pre-visit knowledge) account for differences in students' novelty experience. For example, previous experience with technology-related out-of-school learning opportunities and previous knowledge are linked to a higher curiosity state during the laboratory session. Moreover, tinkerers and students with a more expressed curiosity trait also reported a more positive novelty experience. The only novelty influence factor that was relevant in predicting educational outcomes was technological capability with tinkerers reporting a slightly more positive attitude and higher interest than direction seekers. Likewise, students' novelty experience (measured through cognitive load, curiosity state, exploration behaviour and setting orientation) is linked to their educational outcomes with students' exploration behaviour and 'oriented feeling' playing an important role. In general, the resulting educational outcomes (such as changes in dispositional interests and self-concepts) were mixed but consistently higher for boys.

2.3.4 Novelty influence and experience factors in S'Cool LAB

Out-of-school learning opportunities such as S'Cool LAB are characterised by very high level of novelty, which can influence cognitive and affective learning. Learners' perceived novelty can be beneficial for learning but also overwhelm them. In S'Cool LAB, the novel high-tech experimental equipment and the unfamiliar learning environment require a certain amount of mental capacity. Moreover, particle physics is a difficult learning topic and already provides a high amount of intrinsic cognitive load. Hence, it is crucial to find the right level of novelty. Students' on-site novelty experience can be optimised, for example, by carefully designing student worksheets (enough guidance, limited amount of text and other extraneous load). Furthermore, experimental tasks should be carefully chosen to focus on the activities relevant to conceptual learning rather than technological capabilities, such as connecting cables or calibrating instruments. In the end, different experimental workshops can be compared based on students' perceived novelty experience and adapted if necessary.

[3] https://www.mobillab.ch

To study the effect of novelty, it is crucial to identify important novelty influence factors and their effects on the desired educational outcomes. As novelty is a multi-facet construct, novelty influence factors need to be assessed representing multiple novelty dimensions. In particular, since physics curricula and teaching practices vary between countries, novelty influence factors provide a useful measure to take fundamental differences between different groups of learners into account.

Next, it is important to select novelty experience factors that can have an impact on educational outcomes. Again, multiple facets of novelty experience need to be measured to cover multiple dimensions of the novelty construct. In theory, each of the five novelty dimensions described by Cors et al. (2017) could be represented by specific novelty influence as well as novelty experience factors. Finally, identifying influential novelty factors can help to optimise out-of-school learning opportunities by preparing students in advance or by adapting characteristics of the learning environment.

Table 5 summarises the multi-dimensional novelty factors and educational outcomes assessed within the S'Cool LAB project. The following novelty influence factors are considered: cognitive factors (e.g., pre-visit conceptual understanding, school grades), affective factors (e.g., dispositional physics interest), setting familiarity (previous experience with out-of-school science learning) and capabilities (e.g., previous experience with hands-on experiments). Furthermore, the novelty experience factors include a cognitive factor (cognitive load), setting familiarity (setting orientation), social factors (support by tutors), as well as capabilities (exploration behaviour).

Within the S'Cool LAB project, students' curiosity is considered to be an educational outcome rather than a novelty experience factor (see, for comparison Cors (2016)), first, because only a pre-visit and a post-visit questionnaire are foreseen, which only allows to assess students' retrospectively reported novelty experience in addition to their outcomes, second, because the construct is very similar to interest (see subsection 2.2.3). While Cors (2016) uses prior knowledge and curiosity trait as covariates, all stable student characteristics relevant for the S'Cool LAB project, such as dispositional interest, self-beliefs, and curiosity trait are considered a special type of novelty influence factor in the following. Here it is assumed that the effect of students' prior knowledge (part of the cognitive novelty factor) on the students' cognitive outcomes is not substantially different from the effect of students' dispositional interest and self-beliefs (part of the affective novelty factor) on their affective outcomes.

Table 5. Framework of novelty factors & educational outcomes in the S'Cool LAB study.

Novelty influence factors (NIFs)	Novelty experience factors (NEFs)	Educational Outcomes
Cognitive: School grades Maths, Physics, understanding (pre-visit), experience with particle physics	**Cognitive**: cognitive load	**Cognitive**: understanding (post-visit), cognitive activation, perceived easiness, self-beliefs
Affective: dispositional interest & self-beliefs, curiosity state particle physics, curiosity trait		**Affective**: situational interest, curiosity towards particle physics, perceived enjoyment and helpfulness
Setting familiarity: experience with out-of-school science learning	**Setting familiarity**: setting orientation	
	Social: Tutor support	
Capabilities: experience with hands-on experiments, English skills	**Capabilities**: exploration behaviour	

Conceptualisation and measurement of novelty influence and experience factors

Measuring novelty factors relies on self-reports by students. For example, school grades and experience with particle physics are measured through single items. Cors (2016) introduces items to assess students' prior experience with out-of-school learning opportunities and suggests summarising these experiences based on the desired educational outcomes, for example, within students' out-of-school science learning experience. Börlin (2012) studied the use of experiments across different countries and presented items to characterise students' experience with different types of experiments. The conceptualisation of the remaining novelty influence factors representing dispositional or trait constructs of the educational outcomes (such as interest in physics) is described in the corresponding subsection in section 2.2.

Novelty experience variables, in particular, must be considered a mixture of objective characteristics of the out-of-school learning experience and students' subjective perceptions of the latter. Pawek (2009), for example, introduces a five-level scale to assess students' perception of tutors in out-of-school learning labs. Moreover, Cors (2016) and Molz (2016) use similar Likert scales to assess cognitive load as well as setting orientation and exploration behaviour.

2.4 Students' conceptions of particles in fields, radiation and radioactivity

From a constructivist viewpoint, learners process and integrate new information based on their existing knowledge structures. Or as Ausubel (1968) points out: *"The most important single factor influencing learning is what the learner already knows. Ascertain this and teach [them] accordingly".* Hence, educators need to know about their learners, including their intuitive ideas about the learning content because these might influence or even hinder the learning process. Students' conceptions can be seen as *"learner's internal representations constructed from the external representations of entities"* (Duit & Treagust, 2012). They can originate from prior instruction, intuition, everyday experiences and language, as well as representations in popular science media and they often differ from a scientifically correct explanation (Schecker, Wilhelm, Hopf, & Duit, 2018). While the awareness of students' conceptions is crucial for teachers in formal education institutions (Driver, 1981), it is also important for educators in museums and other out-of-school learning places (Henriksen & Jorde, 2001). Therefore, this section introduces students' conceptions that are relevant for the S'Cool LAB project.

Why focus on electrically charged particles in magnetic fields and radioactivity?

There has been extensive research on students' conceptions for a wide range of science education topics (Duit, 2009). For several of these topics, standardised test instruments, so-called concept inventories, have been developed to easily assess students' conceptions before and after instruction in a quantitative way. However, some topics such as particle physics make quantitative assessment difficult, because students lack the appropriate vocabulary (such as weak interaction, colour charge, or anti-electron-neutrino), or simply have no prior experiences with the topics that would lead to any conceptions. Nevertheless, students typically do have ideas about several concepts closely related to particle physics, such as the behaviour of electrically charged particles in magnetic fields, radiation, the interaction of particles with matter, ionisation, or radioactivity, because these concepts are taught in school or are present in the media. Here, the lack of particle physics knowledge might hinder the learning of these concepts and vice versa.

For example, the behaviour of electrically charged particles in electric and magnetic fields is probably the most important topic with respect to circular particle accelerators, such as the LHC at CERN. Also, atomic and sub-atomic level explanations form the basis of understanding macroscopic phenomena of radioactivity and radiation (Klaassen, Eijkelhof, & Lijnse, 1990; Millar, Klaassen, & Eijkelhof, 1990; Prather & Harrington, 2001). Moreover, learners who understand radioactivity and radiation better evaluate associated risks of

radiation and radioactivity more accurately, express a more positive attitude and less fear (Colclough, Lock, & Soares, 2011; Cooper, Yeo, & Zadnik, 2003; Eijkelhof, 1990). Therefore, understanding radiation and radioactivity, including their sub-atomic explanations, are an important aspect of scientific literacy.

The knowledge of students' conceptions forms the basis of developing learning activities within the S'Cool LAB project. Furthermore, S'Cool LAB is also used to advance our understanding of students' conceptions about selected concepts. Therefore, subsection 2.4.1 summarises students' conceptions about electrically charged particles in magnetic fields and subsection 2.4.2 reviews students' conceptions of radiation and radioactivity in detail.

2.4.1 Students' conceptions of electrically charged particles in magnetic fields

Most of the research on students' conceptions within this topic focuses on university students. For this target group, several suitable concept inventories have been developed to diagnose students' conceptions of electromagnetism, such as

- the 'Conceptual Survey of Electricity and Magnetism' (CSEM) (Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001),
- the 'Brief electricity and magnetism assessment' (BEMA) (Chabay & Sherwood, 1997b; Ding, Chabay, Sherwood, & Beichner, 2006),
- the 'Colorado Upper-Division Electrostatics Diagnostic' (CUE) (Chasteen & Pollock, 2009),
- or the 'Magnetism Conceptual Survey' (MCS) (Li & Singh, 2017).

In general, students only have very limited everyday experience of electrically charged particles in magnetic fields, which makes it difficult to test their conceptions prior to instruction. Maloney et al. (2001) summarise that *"students lack familiarity with both the phenomena and most of the concepts, language, principles, and relations".* Therefore, only a few of the concepts covered by existing concept inventories are also suitable for high-school students.

In many curricula, magnetic fields and magnetic forces are taught after introducing students to electric fields and forces. Therefore, it is no surprise that students tend to confuse the properties of electric fields and magnetic fields. Depending on whether students recently learned about magnetic or electric forces, their recently acquired knowledge negatively interferes with the previously learned material (Scaife & Heckler, 2011).

For example, many students do not differentiate between electric charges and magnetic poles, either by assigning a positive or negative charge to magnetic poles or by expecting an attractive or repulsive electric force caused by the magnetic poles without assigning them an electric charge (Maloney, 1985; Maloney et al., 2001). These conceptions might be caused by students' everyday experience with magnets, which either attract or repel. In fact, all forces students encounter in their everyday life are attractive or repulsive and their direction can be derived from the direction of the respective field. Therefore, it is no surprise that students apply this reasoning in the context of magnetic fields and that these conceptions are very persistent even after instruction (Scaife & Heckler, 2011). Furthermore, students have difficulties differentiating between electric and magnetic field line representations (Maloney, 1985). However, after instruction, items using a field line representation tend to be easier than items using the pole representation of magnetic fields (Scaife & Heckler, 2010). Possibly induced by the selection of typical textbook examples, some students think that electrically charged particles are only deflected in a magnetic field if they move exactly perpendicular to the magnetic field (Bagno & Eylon, 1997). Other students think that electrically charged particles are always affected by a magnetic field, even if they are at rest or move parallel to the field (Maloney et al., 2001). Even if students correctly decide to apply the right-hand-rule, they often predict a wrong force direction, for example, by choosing the direction opposite to the correct direction (Li & Singh, 2017). In addition, students' conceptions from mechanics interfere with magnetism concepts: many students think that objects only move through magnetic fields, because a force acts on them (Li & Singh, 2017).

2.4.2 Students' conceptions of radiation and radioactivity

Various studies investigated learners' conceptions about radiation and radioactivity using a variety of research methods including interviews about scenarios, open & closed questionnaires with different target groups ranging from high-school and university level students, to (pre-service) teachers and the general public (see Colclough (2007) for a detailed overview). In the following, all learners' conceptions that have been identified in either of these studies and samples are listed assuming that conceptions found in research with high-school students might also be relevant for teachers or the general public and vice versa.

Structuring students' conceptions of radiation and radioactivity

Millar et al. (1990) describe the relationships between different scientific concepts associated with ionising radiation (see Figure 5) and suggest a teaching model operating along these concepts at different levels.

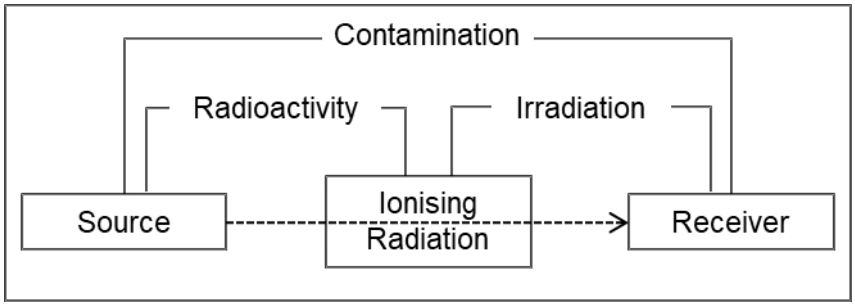


Figure 5. Relationship between scientific concepts associated with radioactivity and ionising radiation (Millar et al., 1990).

Similar to the 'chemistry triplet' (Johnstone, 1991), they describe learning content on macroscopic and (sub-) microscopic levels, though without mentioning a symbolic level. Furthermore, they describe a phenomenological level which connects the concepts to the 'real world experiences' of learners, which has also been adapted by, for example, by Taber (2013). Hence, Millar et al. (1990) suggest starting at a phenomenological level (e.g., X-ray scans in hospitals, experiments with different types of radiation), before continuing with the qualitative and quantitative treatment of the topic at a macroscopic level (e.g., activity and half-life of radioactive sources, radiation dose) and ending with (sub-) microscopic explanations (e.g., atomic nuclei, ionisation, DNA damage).

Indeed, these concepts, relationships and levels can also help to structure students' conceptions. Figure 6 shows an extended framework of radiation and radioactivity concepts at macroscopic and sub-microscopic levels, which has been developed based on (Millar et al., 1990).

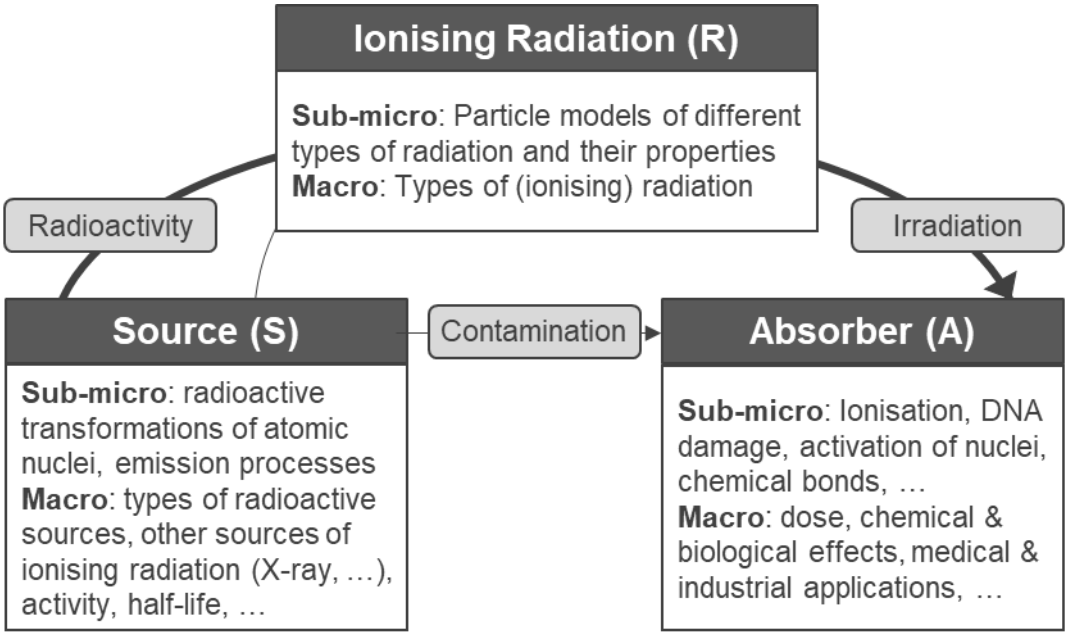


Figure 6. Extended framework of radiation & radioactivity concepts at macroscopic and sub-microscopic levels based on (Millar et al., 1990).

Instead of ‘receiver’ it uses the term ‘absorber’, because this term is typically used in radiation physics (see, for example, vocabulary used by Podgoršak (2014)) and more accurate when focusing on effects of exposure to radiation as it refers more to energy transfer processes. In addition to the description of levels, it includes the concept of ‘activation’ for the absorber. Furthermore, it includes a direct link from ‘source’ to ‘ionising radiation’ to include other sources of ionising radiation, such as X-ray machines or particle accelerators. The following sections describe students’ conceptions and difficulties with respect to the concepts in Figure 6 that have been identified by previous studies.

Use of terminology

Learners typically have problems in using the correct terminology and have problems to distinguish concepts associated with radioactivity and ionising radiation, such as ‘radioactivity‘, ‘radiation‘, ‘radioactive‘, ‘radioactive substance‘ (Eijkelhof, Klaassen, Lijnse, & Scholte, 1990; Eijkelhof, 1990; Henriksen & Jorde, 2001; Maidl & DeKay, 2012; Millar, 1994; Riesch & Westphal, 1975). In particular, learners often confuse the terms (and concepts) ‘contamination’ and ‘irradiation’ (Colclough et al., 2011; Eijkelhof et al., 1990; Eijkelhof, 1990; Millar, 1994; Millar et al., 1990; Prather, 2005). Prather (2005) reports that students use the term ‘unstable’ when talking about radioactive phenomena synonymously with ‘radioactive’. Furthermore, learners confuse activity and dose (Eijkelhof et al., 1990).

Sources of ionising radiation: the sub-microscopic level (S-mi)

To understand the sources of ionising radiation, learners need to understand the different radioactive transformations of atomic nuclei. This requires knowledge of the structure of atoms (nucleus with protons & neutrons, orbital with electrons), the concept of radioactive nuclei, as well as knowledge about natural occurring isotopes. In addition to different transformations of radioactive nuclei (most relevant for the high-school curriculum: alpha, beta, gamma, spontaneous fission), learners need to understand additional emission processes such as bremsstrahlung to understand other the mechanisms of radiation sources (e.g., X-ray machines).

Documented students’ conceptions about **radioactive atoms**:

- *Students show a poor understanding of atomic models* (Johnson & Hafele, 2010; Prather, 2005)
- *The stability of the nucleus depends on the number of electrons in the orbital* (Johnson, 2013; Prather, 2000; Tekin & Nakiboglu, 2006)
- *The stability of the nucleus depends on its density* (Johnson, 2014; Tekin & Nakiboglu, 2006)
- *Unbalanced number of electrons and protons causes radioactivity* (Hafele, 2012)

- *Students confuse ions and isotopes* (Johnson, 2014)

Documented students' conceptions about **radioactive decay**:

- *Radioactive decay involves electrons in the orbital to drop to a lower energy level* (Prather, 2000)
- *Radioactive decay involves electrons being emitted from the orbital* (Prather, 2000)
- *During radioactive beta decay, atomic nuclei split into two parts* (Colclough, 2007)
- *During radioactive decay, the source is disappearing rather than transforming* (Prather, 2000)
- *During radioactive decay, the atom loses pieces/gets smaller until it disintegrates* (Prather, 2005)
- *During radioactive decay, a neutron is emitted by the nucleus* (Prather, 2005)
- *During radioactive decay, differently charged particles are ejected so the overall charge is not balanced anymore* (Prather, 2005)

Sources of ionising radiation: the macroscopic level (S-ma)

Learners need to know about different types of radioactive sources (e.g., certain rocks, radioactive preparations, Radon gas or Uranium glass) and describe their properties, such as their half-life. In addition, they need to know about other sources of ionising radiation such as X-ray machines, particle accelerator, or cosmic sources, such as supernovae.

Documented students' conceptions about **sources of ionising radiation**:

- *Natural (cosmic & terrestrial) and human-made sources of ionising radiation are unknown to students* (Boyes & Stanisstreet, 1994; Colclough, 2007; Prather, 2000; Rego & Peralta, 2006; Riesch & Westphal, 1975)
- *All objects emit radiation because everything is a little bit radioactive* (Klaassen, 1995; Neumann & Hopf, 2013)
- *Radioactive materials are explosive* (Kaczmarek, Bednarek, & Wong, 1987)
- Historical misconceptions about natural radiation, both resolved by experiments by Victor Hess, for which he was awarded the Physics Nobel Prize in 1936 (H. Pleijel, 1936)
    - *The soil is the only source of natural radiation*
    - *Cosmic radiation comes mainly from the sun*

Documented students' conceptions about **half-life**:

- *Half-life is the time in which*
  - *half the radioactive atoms disappear/the number of atoms in total decreases by one half* (Prather, 2000)
  - *half the mass and/or volume of a radioactive substance disappears* (Prather, 2000)
  - *a nucleus loses half of its radiation* (Prather, 2000)
  - *a source is dangerous* (Eijkelhof et al., 1990; Prather, 2000)
- *Half-life depends on physical conditions, such as temperature, pressure, chemical condition* (Acar Sesen & Ince, 2010; Tekin & Nakiboglu, 2006)
- *Radioactive sources remain active for a very long time* (Alsop, 2001)
- *Half-life is the life-time of an individual nucleus that decays continuously (Jansky, 2019)*
- *Students struggle to accept the randomness of the radioactive decay, in particular, because they are unfamiliar with continuous random variables, such as 'time' (Jansky, 2019)*

Ionising radiation: the sub-microscopic level (I-mi)

This category includes knowledge about particle (and wave) models of different types of ionising radiation.

Documented students' conceptions about **ionising radiation**:

- *Ionising radiation is described as material* (Lijnse, Eijkelhof, Klaassen, & Scholte, 1990; Maidl & DeKay, 2012), *waves, energy, light/rays, process of particles leaving an object, field, force, heat, current* (Johnson, 2014), *some kind of gas* (Eijkelhof, 1990), *or immaterial entity* (Alsop, 2001)
- *Gamma radiation consists of waves, other types of radiation travel in a straight line* (Neumann & Hopf, 2013)
- *Radiation consist of radiating/radioactive particles* (Millar & Gill, 1996; Neumann & Hopf, 2012; Prather, 2000)

Ionising radiation: the macroscopic level (I-ma)

Many students are unaware of different types of ionising radiation and their unique properties (Colclough, 2007; Johnson, 2014). Instead, they list radiation from computers, TVs, microwaves, batteries, cell phones, lasers, as well as radio waves, heat, IR light, visible light, sound, ultrasound, and Wi-Fi as ionising radiation (Acar Sesen & Ince, 2010; Eijkelhof, 1990; Johnson & Maidl, 2014).

Furthermore, their conceptions of the sub-microscopic nature of ionising radiation affect their beliefs about the propagation of ionising radiation.

Documented students' conceptions about the **propagation of radiation**:

- *Radiation needs a medium such as air to spread* (Eijkelhof et al., 1990; Riesch & Westphal, 1975)
- *Radiation can be stopped by vacuum* (Eijkelhof et al., 1990)
- *Radiation can be stopped by counter-radiation* (Eijkelhof et al., 1990)
- *Propagation of radiation can be influenced by ventilation/wind* (Eijkelhof, 1990)
- *Radiation travels as a cloud* (Eijkelhof et al., 1990)
- *Radiation travels as a physical stream and flows around objects* (Eijkelhof et al., 1990; Riesch & Westphal, 1975)
- *Radiation lingers for a while even without source* (Eijkelhof et al., 1990)

Absorber: Interactions of ionising radiation with matter at the sub-microscopic level (A-mi)

The possible and most likely interactions between ionising radiation and matter depend on the type of radiation and their energy. Typically, these interactions are described on a particle level and distinguished based on whether the particles interact with orbitals or nuclei of the absorber atoms. The most important energy transfer processes for different types of particles according to Podgoršak (2014) are listed below.

- **Photons** (X-ray & γ radiation): Photoelectric effect, Compton scattering, pair production, approx. above 10 MeV: photonuclear reactions (including photon activation)
- **Electrons & positrons** (β^- & β^+ radiation): Coulomb interaction leading to inelastic scattering with orbital electrons of the absorber (including ionisation) and inelastic scattering with nuclei of the absorber (bremsstrahlung)
- **Positrons** (β^+ radiation): annihilation
- **Helium nuclei** (α radiation): Coulomb interaction leading to inelastic scattering with orbital electrons of the absorber (including ionisation), negligible for energy transfer: inelastic scattering with nuclei of the absorber (nuclear reactions)
- **Neutrons** (seldom relevant for high-school curricula): inelastic scattering, neutron absorption including neutron activation

Students show persistent alternative conceptions about absorption of radiation. Hence, Millar and Gill (1996) recommend that *"the absorption process should be discussed more explicitly in teaching programmes"* and more research on students' conceptions of

absorption models and their limitations is needed, for example, by comparing students' ideas about the absorption of light with their ideas about the absorption of ionising radiation. So far, previous studies list only a small number of students' conceptions related to the sub-microscopic absorption of ionising radiation.

Documented students' conceptions about **ionisation**:

Maidl and DeKay (2012) state that *"Ionization by ionizing radiation is the key phenomenon that must be understood to make sense of the effects of radiation."* However, students demonstrate a poor understanding of atomic models, ions and ionisation (Hafele, 2012; Johnson, 2014), and therefore confuse ionisation and radioactive transformation processes, see section 'Sub-microscopic level (S-mi)' (Colclough, 2007; Hafele, 2012).

Additional documented students' conceptions of the **interaction of radiation with matter**:

- 'Internal blocking' picture: *radiation either hits particles inside an object and is thereby stopped or passes through gaps between particles* (Colclough, 2007)
- *Students show a limited understanding of the photoelectric effect* (McKagan, Handley, Perkins, & Wieman, 2009; Oh, 2011; Taslidere, 2016)
- *Learner confuse properties X-rays and light, in particular, the transparency/opacity of materials to different radiation types* (Clement & Fisseux, 1999)

Absorber: Interactions of ionising radiation with matter at the macroscopic level (A-ma)

Interactions of ionising radiation with matter rarely leads to any macroscopic effects. Typically, radiation doses are far too small to cause measurable effects, for example, in temperature or deterministic biological effects. However, many students think that radiation accumulates inside objects "*like the storage of water by a sponge*" (Alsop, 2001). Often this is combined with a conception of 'conservation of radiation'. Furthermore, they show inaccurate conceptions of macroscopic effects of radiation differentiating in particular between living and non-living things.

Documented students' conceptions of **accumulation & conservation**:

- *Radiation accumulates in front of objects* (Riesch & Westphal, 1975)
- *Radiation accumulates inside objects* (Alsop, 2001; Eijkelhof et al., 1990; Riesch & Westphal, 1975)
- *Radiation is conserved, if an object absorbs radiation it will later release the radiation* (Eijkelhof, 1990)
- *Radiation attaches itself to objects* (Eijkelhof et al., 1990)

Documented students' conceptions about other **macroscopic effects of radiation**:

- *Radiation causes objects to get warm or to smell* (Colclough, 2007; Riesch & Westphal, 1975)
- *Irradiated objects start to glow* (Boyes & Stanisstreet, 1994)
- *Radiation is either stopped completely or not at all* (Eijkelhof et al., 1990)
- *Radiation is reflected by objects* (Colclough, 2007; Eijkelhof et al., 1990; Riesch & Westphal, 1975), *reflected by bones* (Plotz, 2017), *and refracted by water* (Colclough, 2007)
- *Metal objects are affected more by radiation than non-metal objects because metals attract radioactivity* (Alsop, 2001)

Documented students' conceptions about **biological effects of radiation**:

- *Reasons for biological harmfulness are unknown, students think radiation is poisonous* (Lijnse et al., 1990; Riesch & Westphal, 1975)
- *Radioactivity invades genes like a virus invades a cell* (Alsop, 2001)
- *Humans can feel radiation as stitches* (Riesch & Westphal, 1975)
- *Radiation is painful* (Klaassen, 1995)
- *The life span of radiologists is shortened (Kaczmarek et al., 1987; Mubeen, Abbas, & Nisar, 2008)*
- *Human bodies have a defence system against radiation, as long as it is not too much* (Eijkelhof, 1990)
- *Radiation standards indicate a safety level of radiation dose* (Eijkelhof et al., 1990)
- *Radioactive substances are always dangerous* (Eijkelhof et al., 1990)
- *Artificial radiation is more dangerous than natural radiation* (Eijkelhof et al., 1990)
- *Radiation leads to impot*ence (Eijkelhof et al., 1990)
- *Living things in particular are affected by radiation, they "actively attract and soak up" radiation* (Alsop, 2001)
- *Radiation sickness is the immediate effect of irradiation* (Alsop, 2001)
- *Plastic suits protect workers from radiation* (Alsop, 2001)
- *Radiation will spread from one body part to the rest of the body* (Johnson & Maidl, 2014)
- *Radiation remains in an object/body for years* (Riesch & Westphal, 1975)
- *Gamma rays are more dangerous than X-rays* (Mubeen et al., 2008)

Contamination versus irradiation (versus activation)

Probably the most problematic students' conceptions are related to the distinction between the concepts of contamination, irradiation and activation (Colclough et al., 2011; Eijkelhof, 1990; Kaczmarek et al., 1987; Millar et al., 1990). Many students believe that exposing objects to radiation (irradiation) makes these objects radioactive themselves:

- *Irradiated material becomes radioactive* (Colclough, 2007; Henriksen & Jorde, 2001; Riesch & Westphal, 1975)
- *Irradiation of food is dangerous because the food will contain radiation / become radioactive* (Eijkelhof et al., 1990; Johnson, 2014)
- *Length of exposure determines how radioactive objects get* (Prather, 2000)
- *Size of absorber matters, the bigger the more radioactive they get* (Millar, 1994)
- *After an X-ray scan, objects and human bodies emit radiation* (Eijkelhof, 1990; Kaczmarek et al., 1987; Mubeen et al., 2008)
- *Confusion between 'containing radioactive matter', 'containing radiation', 'being radioactive'* (Millar, 1994)
- *Heating or chemical treatment will treat contamination* (Eijkelhof et al., 1990)
- *Certain chemicals can make normal matter radioactive* (Johnson & Maidl, 2014)

However, students seem to apply different reasoning patterns linked to their respective conception of ionising radiation. Especially students who do not distinguish between radiation and radioactivity seem to describe radiation as material (as some kind of contaminant) and belief exposure will make things radioactive (Hafele, 2012; Millar & Gill, 1996). Attributing matter-like properties to radiation is also linked to the conception of 'radiation conservation', which leads students to believe that radiation accumulates inside objects or is contained inside objects and is released later (Eijkelhof, 1990; Millar & Gill, 1996). Furthermore, some students think of radiation as radiating (or radioactive) particles (Millar et al., 1990; Neumann & Hopf, 2012) that can be transferred to another object similar to what happens during contamination. In addition, students might consider that radiation activates the atomic nuclei of the absorber. Although this process is possible, for example, with neutron radiation or very high photon energies (approx. $> 10\ MeV$), most radioactive sources will not produce radiation that can trigger activation. Also, Molz (2016) found that the concept of activation through radiation is unknown to most students.

Attitudes towards radiation and radioactivity

Compared to students' cognitive conceptions about radiation and radioactivity, relatively little is known about their attitude towards radiation and radioactivity. This is a challenge as both are closely related and important in risk evaluations and decision making (Colclough,

2007). Students typically fear radiation (Cooper et al., 2003; Eijkelhof et al., 1990; Eijkelhof, 1990; Neumann & Hopf, 2012), for example, because they perceive it as dangerous or harmful especially for living organisms (Acar Sesen & Ince, 2010; Henriksen & Jorde, 2001; Klaassen, 1995). Students associate radiation and radioactivity with pollution (Boyes & Stanisstreet, 1994) or poison (Lijnse et al., 1990).

However, students' emotions towards ionising radiation seem to depend on the context. Some students perceive radiation as safe in hospitals (Eijkelhof, 1990; Millar, 1994; Neumann & Hopf, 2012), often while underestimating the risk of medical imaging (Klaassen, 1995; O'Sullivan et al., 2010). Whereas some students fear in particular artificial radiation more than natural radiation (Eijkelhof et al., 1990; Plotz & Hopf, 2016), others perceive all radiation as unnatural but distinguish between good and bad radiation (Henriksen & Jorde, 2001). Here, Colclough (2007) developed a questionnaire with 20 statements and Likert-type response scales to assess pre-service teachers attitude towards radioactivity, for example, *"I would be scared to perform […] experimental demonstrations using school radioactive sources."* or *"I regard radioactivity as an emotional subject."*.

3 Research gaps and research questions

As discussed in the previous chapter, out-of-school learning offers unique opportunities for learning and development with a broad range of intertwined educational outcomes. Furthermore, learning and development are majorly affected by students' novelty experience, which in turn is affected by students' novelty influence factors. However, several open questions about out-of-school learning remain. For example, how can educators make use of their potential for cognitive learning, or as Renninger (2007) asks *"What conditions support participants to shift from exploring science content to posing questions that characterize science literacy?"*. Even though we already know a lot about some of the possible educational outcomes (such as self-beliefs or interest) from formal learning settings, studying them in an out-of-school learning setting is useful since *"The inquiry-oriented, free-choice ISL [Informal Science Learning] setting offers the possibility of reexamining what has been understood about these variables and their relation to learning."* (Renninger, 2007).

Assessing students' conceptual understanding before and after a visit to an out-of-school learning lab relies on suitable test instruments. However, a concept test for particle physics does not exist (yet); only a few items from previous concept tests are suitable to capture isolated aspects of high-school students' understanding. Actually, the unknown vocabulary and lack of (every-day) experience make it extremely challenging to assess students' conceptions of this topic. However, related topics such as ionising radiation and radioactivity are based on conceptual understanding of elementary particles and their interaction with other particles.

3.1 Limitations of previous studies

Although several studies examined the educational effectiveness of out-of-school learning laboratories, their results are mixed and their generalisation limited (see subsection 2.1.5).

First, some of these studies are based on relatively simple hands-on activities that can be performed at school without any problems. In this case, the out-of-school learning environment serves mostly as a decorative element with a high amount of unjustified novelty and cognitive load without exploiting its unique potential. Therefore, it is no surprise that the effect of these out-of-school learning opportunities appears limited compared to a formal schooling setting. Advanced-level particle physics experiments with high-tech equipment that is usually not available to schools might have better chances to impact learners than, for example, experiments focusing on the chemistry of starch.

Second, most studies on out-of-school learning labs took place in Germany with a German sample while using single-level models on nested data. Students from different cultural and educational backgrounds might react differently to hands-on experimentation in out-of-school settings. Even students from different schools might bring different experiences that could be reflected in their educational outcomes.

Third, the effects of novelty factors played mostly a side effect in previous studies and were not assessed systematically or multi-dimensionally. Hence, we are only beginning to systematically measure and understand their unique role in out-of-school learning opportunities.

Last, most studies investigating cognitive effects of out-of-school learning labs measured only students' factual knowledge gain. However, the power of interest (and potentially other affective variables) seems to become manifest in learning only when measuring cognitive outcomes requiring deeper level processing.

Besides these open questions of science education research, Renninger (2007) also highlights the relevance of formative assessment for out-of-school learning places: *"One issue, of course, is that ISL research is not only research on interest and motivation in ISL, but also needs to be use-informed (Stokes, 1997). Findings from research that is undertaken need to provide information to those who will use it to inform practice."*

The existing research on students' conceptions covers a wide range of science education topics. Several studies investigated learners' conceptions about radiation and radioactivity using a variety of research methods and different target groups. But only a few studies report quantitative results of larger sample sizes allowing statistical analyses, such as factor analysis or testing for reasoning patterns. Furthermore, most of the previous studies in this field are comparatively old (reaching back to 1975) and it is not clear, which of the previously documented conceptions are still relevant for today's learners. Also, we are lacking a structure of associated concepts and conceptions to understand students' reasoning patterns. Instead, previous research only reports conceptions related to single isolated concepts, such as half-life. Here, item development and more research is needed to paint a bigger picture of students' conceptual understanding of this topic and to assess students' conceptions in a time-efficient way.

3.2 Motivation for this research project

Previous results suggest that out-of-school learning labs need to use their unique potential with respect to special features: field trips, access to rare equipment, hands-on experimentation, and contact to scientists. However, the effectiveness of the activities depends on their design and their potential to engage students in cognitively activating activities while fostering their interest and self-beliefs.

The first steps of the S'Cool LAB project included the evidence-guided development of high-quality, hands-on activities. In this way, the S'Cool LAB full-day programme is expected to provide an extreme example of an out-of-school learning lab with the best possible conditions for learning and development (see chapter 4 for a brief overview of this process).

The S'Cool LAB evaluation study aims to prove the educational effectiveness of S'Cool LAB's hands-on sessions with an international sample of high-school students while taking the nested data into account. Furthermore, the role of student characteristics (or novelty influence factors) and students' perception of laboratory characteristics (or novelty experience factors) are studied systematically to gain inside into their predictive power for a wide range of educational outcomes and the 'mechanism of action' of out-of-school learning labs.

Furthermore, participants of S'Cool LAB's full-day programme are used to expand our current understanding of students' conceptions of ionising radiation and radioactivity as well as students' reasoning patterns while developing and validating a suitable quantitative test instrument.

3.3 Research questions

3.3.1 Affective & cognitive effects and outcomes

Although many of the educational outcomes consist of both affective and cognitive aspects, and distinguishing cognition and affect is somewhat arbitrary, the research questions are separated into affective and cognitive effects in the following to align with the theoretical framework of the concept map presented in subsection 2.2.7. For example, self-beliefs are considered here a more cognitive construct, although they are often grouped with interest as motivational variables.

- **RQ1.1)** To which extent can half-day hands-on sessions in S'Cool LAB foster **students' affective variables** a) physics interest, b) curiosity towards particle physics? (subsection 7.4.5)
- **RQ1.2)** To which extent can half-day hands-on sessions in S'Cool LAB foster **students' cognitive variables** a) physics self-beliefs, and b) conceptual understanding? (subsection 7.4.5)
- **RQ2)** Which of the student and laboratory characteristics are relevant for modelling the **affective educational outcomes** a) situational interest, b) curiosity towards particle physics, c) enjoyment, and d) helpfulness and the **cognitive educational outcomes** e) situational self-beliefs, f) perceived easiness, g) cognitive activation, and h) conceptual understanding of S'Cool LAB? (subsection 7.9.1)
- **RQ3)** To which extent do female and male participants differ in their educational outcomes? (subsection 7.4.5)
- **RQ4)** To which extent do educational outcomes differ among students coming in different groups, from different teachers, and from different countries? (subsection 7.9.1)

3.3.2 Additional questions as part of the formative assessment of S'Cool LAB

Besides S'Cool LAB's educational outcomes and novelty factors, the S'Cool LAB evaluation study also aims to provide information for CERN and the S'Cool LAB team, which can be used to improve CERN's educational offers (section 7.6).

- **Q1)** How do students perceive the hands-on session in S'Cool LAB compared to their guided tours at CERN?
- **Q2)** Which of the workshops in S'Cool LAB do students like most?
- **Q3)** Which of the workshop is associated with a high amount of cognitive load?
- **Q4)** How do students rate the importance of different aspects of their trip to CERN?

3.3.3 Students' conceptions of ionising radiation and radioactivity

The following research questions guide the development and evaluation of items assessing students' conceptions of ionising radiation and radioactivity (section 8.8).

- **RQ5.1)** Which of the students' conceptions identified by previous studies are relevant to today's learners taking part in S'Cool LAB workshops?
- **RQ5.2)** What are the similarities and differences between a hypothetical two-level source-radiation-absorber model and the factor structure of the developed Multiple-True-False items?
- **RQ5.3)** What are common reasoning patterns exhibited by clusters of students?

4 Design of S'Cool LAB and its learning activities

CERN, the world's largest particle physics laboratory, has a longstanding tradition to science communication to share its research results with the general public. In 2018, more than 100 000 visitors from more than 60 different countries came to CERN in the framework of free-of-charge guided tours of CERN's research facilities. More than half of them were high-school students aged 16 to 19. Their teachers usually plan school trips to CERN to show their students a prime example of modern physics research, where physicists from around the world work together to find answers to the fundamental questions of humankind. Teacher hope that the fascination of gigantic machines, huge collaborations, and tiny particles inspires their students and allows them to perceive physics as a highly active field of research. Finally, teachers use CERN as an out-of-school learning opportunity because they sometimes lack the equipment or knowledge to teach particle physics adequately. To offer more learning activities at CERN, especially for the future generation of scientist and decision-makers, S'Cool LAB was founded as a hands-on learning laboratory for high-school students. This chapter provides an overview of the history and goals of S'Cool LAB (section 4.1) as well as its educational offers (section 4.2). After highlighting selected aspects of the design of its learning activities (section 4.3), section 4.4 introduces the S'Cool LAB Day programme, which was used to collect data from students in S'Cool LAB. Finally, section 4.5 describes the three workshops that formed the hands-on S'Cool LAB session, which was the focus of the studies presented in chapters 6 (pilot study) and 7 (main study).

4.1 History and goals of S'Cool LAB

Plans for a dedicated hands-on lab for high-school students at CERN existed for several years. However, only in 2013, when CERN received a generous donation from the 'Lotterie Romande', it was decided to move forward with the S'Cool LAB project. Between March 2013 and June 2014, a 200 m^2 space was separated from CERN's existing Microcosm exhibition and reconstructed as a laboratory space for high-school students. The first step of this PhD project, which started in March 2014, was to develop hands-on activities in an iterative design process, to put the lab into operation, and to establish hands-on workshops in S'Cool LAB as a testbed for physics education research. This process included extensive equipment orders and safety documentation (for working with lasers, cryogenics, electrical installations, radioactive sources), establishing stable work conditions, website content development, planning and implementation of a web booking tool, and building up and managing a diverse team of up to 8 team members and more than 50 volunteers from

CERN's scientific community. Until today, more than 34 000 participants took part in particle physics workshops in S'Cool LAB. Approximately 95% of the student groups come from CERN's member and associated member states, but students also arrive from countries further away, such as Australia, the US or Brazil (see Figure 7).

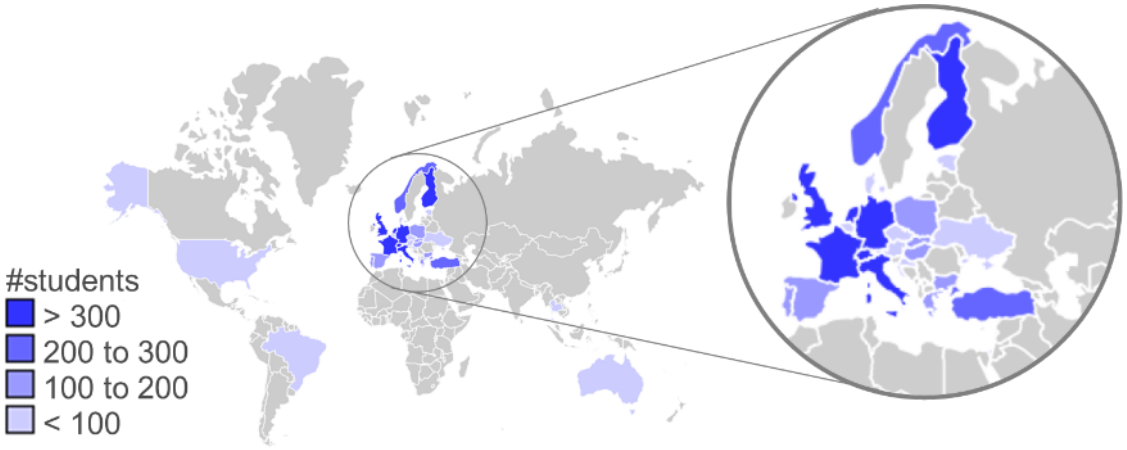


Figure 7. Origin of student groups taking part in S'Cool LAB workshops in 2017.

The workshops in S'Cool LAB aim to a) give insights into the working methods, technologies and research of the world's largest particle physics laboratory, b) spread CERN's spirit of science curiosity, foster physics interest and self-beliefs, and c) to make CERN's physics and technologies understandable to students through hands-on experimentation.

In particular, students in S'Cool LAB have the chance to experience concepts of particle accelerators and detection themselves in a hands-on way supported by volunteers from CERN's scientific community. Here, S'Cool LAB tries to offer enjoyable experiences with tasks that allow exploration but do not overwhelm students as recommended by Allen (2004) and Martin (2004). Furthermore, S'Cool LAB serves as CERN's physics education research testbed, bringing together students and their teacher from around the world with a team of physics education researcher, mediated by hands-on minds-on particle physics experiments. A Feynman-like diagram summarises this aspect in Figure 8.

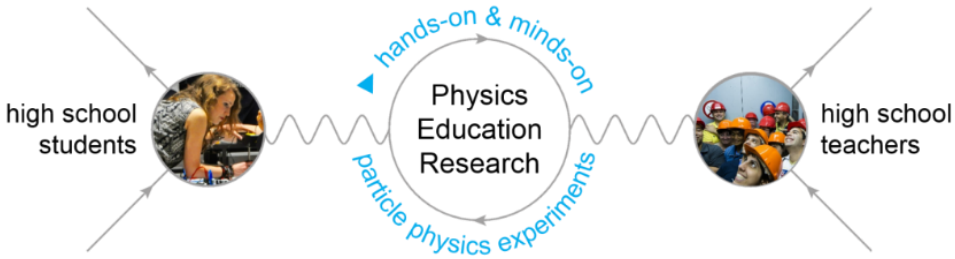


Figure 8. S'Cool LAB, CERN's hands-on particle physics learning lab & testbed for education research.

This research project in particular, aims to study the 'coupling' of students to these hands-on experiments to understand their cognitive and affective effects on students as well S'Cool LAB's 'mechanism of action' better. Research results are used to continuously improve S'Cool LAB's offers and to contribute to the understanding of out-of-school learning places in general.

4.2 S'Cool LAB's educational offers

S'Cool LAB's experimental activities have been developed in an iterative process. All 90-minute workshops focus on one aspect of CERN, such as particle detection, particle acceleration, medical applications of particle physics, or key technologies, such as superconductivity or particle traps. Students work in teams of three to four guided by process worksheets and supported by volunteering CERN scientists (so-called tutors).

The most popular workshop in S'Cool LAB is the 90-minute stand-alone cloud chamber workshop with 5420 student participants in 2018 (Table 6). In addition, every year approximately 700 students take part S'Cool LAB's research programme (so-called 'S'Cool LAB Day programme' or 'S'Cool LAB PLUS+ programme'[4]) which allows them to spend more time on two to three hands-on workshops. In turn, participants take part in CERN's physics education research projects by either filling out questionnaires, being video-recorded while doing experiments, or taking part in interviews. The 'S'Cool LAB Day programme' was also used to generate data for the studies of this PhD project, see subsection 4.4 for a detailed description of this programme.

Table 6. S'Cool LAB participants from 2014 to 2019 taking part in different educational offers.

Educational offer	2014	2015	2016	2017	2018	2019	Total
S'Cool LAB Days / PLUS+ programme	0	430	1030	1030	640	660	**3790**
Cloud Chamber Workshops (students)	720	2420	3370	4800	5420	5820	**22550**
Other (teachers, VIPs, masterclasses,…)	1100	1020	1430	1430	1450	1380	**7810**
Total	**1810**	**3870**	**5830**	**7260**	**7510**	**7860**	**34140**

S'Cool LAB also welcomes approximately 900 teachers per year who take part in professional development programmes at CERN (cern.ch/teachers). In addition to

[4] After the data collection for this PhD project in the framework of the 'S'Cool LAB Day programme' was concluded, the structure of the programme was altered to take preliminary research findings into account and to accommodate other types of physics education research projects. Therefore, the programme has been rebranded to 'S'Cool LAB PLUS+ programme'.

activities on-site, a collection of low-cost hands-on activities has been developed over the past years that can be used by teachers in their own classrooms to introduce certain topics of particle physics without coming to CERN (cern.ch/scool-lab/classroom-activities).

4.3 Design principles for the development of hands-on activities

Since the detailed description of the development process of all hands-on activities goes beyond the scope of this thesis, this section provides a brief overview of important design principles. In particular, S'Cool LAB workshops have been developed following an iterative design process based on the Model of Educational Reconstruction (Duit, Gropengiesser, Kattmann, Komorek, & Parchmann, 2012). Therefore, as a first step, particle physics topics suitable for hands-on experimentation have been identified based on their relevance at CERN and the availability of equipment. The physics contents of these topics were analysed with respect to their elementary ideas.

Second, students' and teachers' perspectives were analysed and integrated. Due to the organisational effort of physics field trips to Switzerland, students taking part in trips to CERN are typically well prepared and motivated to learn more about particle physics. Indeed, teachers prepare their students in their classroom, sometimes by adjusting the order of their physics curriculum, or in the framework of extracurricular activities, such as science clubs. Furthermore, they usually chose experiments in S'Cool LAB based on their individual and their students' preferences as well as their physics curriculum. Teachers are also encouraged to prepare their students for S'Cool LAB with the help of dedicated online material including detailed descriptions of essential key concepts as well as learning material for these key concepts (see section 4.4).

Students' conceptions can influence the learning process of hands-on activities. In particular, some students tend to observe outcomes of experiments differently (i.e., according to their expectations), when holding certain alternative conceptions (Hynd, McWhorter, Phares, & Suttles, 1994; Miller et al., 2013). However, this can also form the basis of cognitively activating experiments with surprising observations that might contradict existing conceptions. Therefore, in addition to literature research about previously documented students' conceptions, student worksheets have been evaluated in the early development stages of the workshops to analyse students' predictions, observations, explanations and other notes.

Furthermore, student groups were observed and interviewed while piloting the experimental workshops between September 2014 and December 2015. In this way, tasks

that were too difficult, too easy, or not clear enough have been adapted, the order of experimental tasks has been modified to support learning, and instructions for handling the equipment have been simplified as much as possible. In addition, the amount and complexity of all texts as well as the number of scientific terms in the student worksheets have been reduced as much as possible, because all students have to follow instructions in English independent of their nationality.

The complexity, cost and safety features of the research equipment in S'Cool LAB seldom allows free exploration without guidance. However, when possible, several optional tasks were included to give students the freedom to choose tasks according to their preferences and abilities.

4.3.1 Predict-Observe-Explain (POE)

Student worksheets in S'Cool LAB are structured by using Predict-Observe-Explain tasks (see Figure 9) where possible. White and Gunstone (1992) introduced this type of learning/teaching sequence to diagnose students' understanding of science concepts. Furthermore, POE tasks provide students with cognitive conflict, and as also Vera Rubin suggests *"Science progresses best when observations force us to alter our preconceptions."*

In S'Cool LAB, students first **predict** the outcome of an experimental manipulation based on their previous knowledge or their intuition. This step helps them activate important underlying physics principle but also allows tutors to diagnose students' conceptions. As recommended by White and Gunstone (1992) and to prevent social loafing, every student is asked to provide their own predictions. To reduce the time spent on writing, students chose their prediction for some tasks from different answer options or provide an alternative prediction. In most tasks, they are also asked to describe the reasoning that led them to their prediction.

Afterwards, students carefully **observe** the outcome of their manipulation task while focusing on the aspects described in the prediction. While developing the workshops for the hands-on session in S'Cool LAB, students' notes were evaluated with respect to the proportion of correct predictions and correct observations. Indeed, up to 28% of the students observed the outcome of an experimental manipulation wrongly. Specifically, effects that were hard to observe, such as contrast differences on a fluorescent screen in an X-ray machine, were often not observed as intended. Instead, students' observations were often consistent with their incorrect predictions.

Last, students interpret their observations and **explain** deviations from their predictions. In this step, students discuss among their peers but also with the volunteering CERN scientists. The tutors receive information about typical students' conceptions in dedicated training sessions in advance and are thus trained to help students resolve possible cognitive conflicts, which ideally supports conceptual progression.

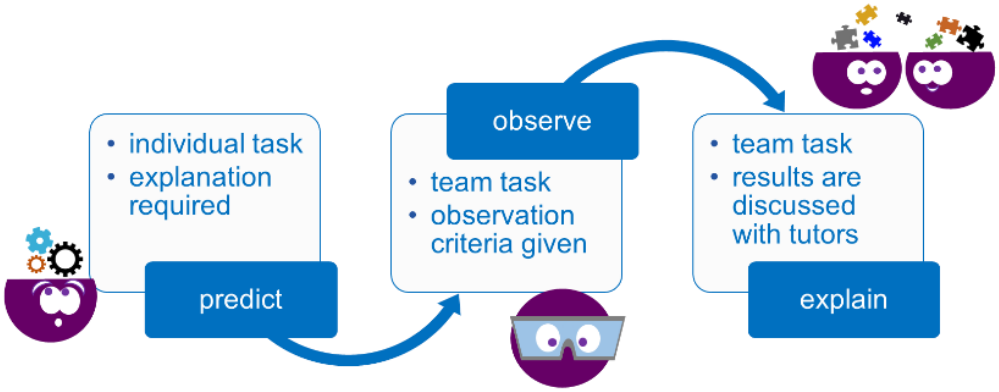


Figure 9. Learning activities in S'Cool LAB are based on known students' conceptions and include Predict-Observe-Explain tasks (White & Gunstone, 1992) to foster conceptual progression.

Before students start doing experiments, a S'Cool LAB team member introduces them to these Predict-Observe-Explain tasks by comparing it to the collaboration between theorists and experimentalists at CERN: Physicists in CERN's theory department develop theories and predict what to observe in certain experiments if these theories were correct, for example, when calculating the probabilities of specific signatures in the LHC detectors. Similarly, students will predict the outcomes of their experiments in S'Cool LAB. Experimentalists, on the other hand, test theoretical predictions by designing, performing and analysing experiments. Similarly, students in S'Cool LAB perform and closely observe their experiments. Contrary to students' common opinion, physicists at CERN are most excited if the predictions differ from the experimental observations because this provides the opportunity for theorists and experimentalists to work together to extend our existing theories. Instead, students are typically afraid to write down 'wrong' prediction and hence need some time to get used to POE tasks and support when explaining their observations.

POE example from S'Cool LAB

In the electron tube workshop, students deflect a low-energy electron beam with a bar magnet. Before they receive the magnet, every student predicts the position of the deflected beam spot for different hypothetical magnet positions (see Figure 10). Typically, more than half of the students expect that the magnet attracts (or repels) the electrons. After the prediction, students hold a magnet as indicated close to the electron beam and note down their observations. Afterwards, the worksheet guides them to summarise their results and to explain the deflection through the Lorentz force and the right-hand-rule.

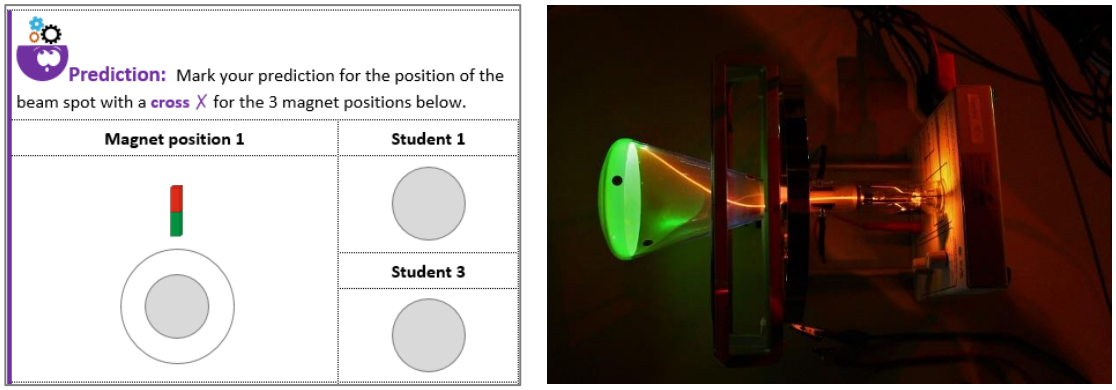


Figure 10. Example of a POE task in the electron tube workshop. On the left: Students are asked to predict the position of the beam spot when holding a magnet close to the beam as instructed. On the right: When students hold a magnet as indicated in the drawing on the left, they will notice that the electron beam is deflected to the side.

Why POE tasks?

Students' conceptions can influence how students observe experiments and one cannot assume that all learners perceive and interpret the outcome of an experiment as expected (Hynd et al., 1994). Miller et al. (2013) studied how university students perceive demonstrations in physics lectures. In their study, 18% of the observations differed from the actual outcome. However, a correct observation was crucial for conceptual learning in their study. Furthermore, students predicting the outcome first, independent of whether the prediction was correct or wrong, were significantly more likely to report the outcome of a demonstration correctly. Similarly, Crouch, Fagen, Callan, and Mazur (2004) report improved learning from lecture-demonstrations if students predict the outcomes first through a multiple-choice question. Results of a study by Brod and Breitwieser (2019) with adults suggest that predicting an answer beforehand stimulates curiosity because the prediction enhances the relevance of the perceived knowledge gap. In their study, high curiiosity was also associated with better memory of the correct answer.

However, students react differently to cognitive conflicts. Trumper (1997) describes different reactions to provocative questions and situations. For example, some students do not recognise the conflict, while other students try to resolve their cognitive conflict using their original conceptions. Niaz (1995) also reports that some students tend to ignore a cognitive conflict and instead protect their existing conceptions by ignoring their observations. Indeed, results of a study with 9-12 year old children by Brod, Breitwieser, Hasselhorn, and Bunge (2019) suggest that predictions only support learning for students with sufficient inhibitory control: When confronted with cognitive conflict, students need to be able to suppress prior beliefs to allow establishing new beliefs.

Last but not least, using predict-observe-explain allows to implement many of the suggestions to improve hands-on activities and to increase their educational effectiveness listed in subsection 2.1.6.

4.4 Description of the S'Cool LAB Day programme

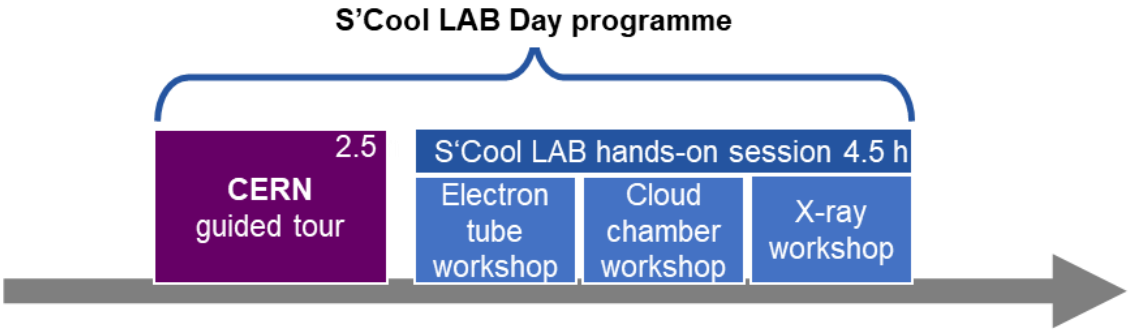


Figure 11. Overview of the S'Cool LAB Day programme at CERN.

Students participating in the studies presented in chapters 6 and 7 took part in a full-day programme at CERN, the so-called 'S'Cool LAB Day programme'. This programme was developed for groups consisting of 12 to 36 students aged 16 to 19 years. After a 2.5-hour guided tour at CERN, students took part in a 4.5-hour hands-on session in S'Cool LAB, which consisted of three 90-minute workshops (see Figure 11).

For this programme, teachers applied on behalf of their student group and were selected based on their motivation and their plans to embed the CERN-experience into their curriculum. Once at CERN, students were introduced to the schedule of the day (see Figure 12) before leaving on a guided tour of two facilities at CERN.

Figure 12. Typical schedule of a S'Cool LAB Day at CERN.

These facilities were chosen to represent topics from particle acceleration and particle detection. For comparability, the same facilities were visited by all groups participating in the S'Cool LAB studies. During the guided tour, students were accompanied by a volunteering CERN scientist who explained the facilities and answered students' questions.

The first visit point was the **synchrocyclotron**, CERN's first particle accelerator, which went into operation in 1957 and contributed majorly to today's understanding of the weak interaction (Fazzini, Fidecaro, Merrison, Paul, & Tollestrup, 1958). After its shutdown and decommission in 1990, it was opened as a modern visit point in 2014. When entering the facility hall, visitors see a timeline with the development of CERN and its research facilities starting in the 1950ies. Upon entering the accelerator area, a video about the history of CERN and the synchrocyclotron is shown. Then, a projection mapping show starts illuminating the front of the accelerator itself explaining its functioning step by step while highlighting the relevant parts. In particular, strong electric fields accelerated protons between two D-shaped electrodes, while strong electromagnets were used to force the protons on circular paths. The radius of the particles' trajectories increased with their energy before they were extracted and directed towards a target.

The second visit point was the **ATLAS visitor centre**. ATLAS is the largest LHC detector, 50 m in length, 25 m in diameter, with a mass of 7000 kg. ATLAS consists of several layers of sub-detectors arranged around the collision point in the centre of the detector and was designed to accurately measure the trajectories and properties of the particles produced in particle collisions in the LHC (Aad et al., 2008). The ATLAS visitor centre is located next to the ATLAS control room. During their guided tour, visitors watch a 3D movie of the construction and operation of the detector. Also, guides showcase the ATLAS control room, where physicists monitor the different components of the detector with the help of numerous computer screens. Guides use an interactive screen to illustrate the layout of the detector and its onion-like structure, demonstrate the unique signatures of different particle types in the detector, and explain examples of collision events in the detector, such as transformations of Higgs boson candidates.

After the guided tour, students returned to S'Cool LAB and were introduced to the laboratory's safety rules and emergency procedure as well as the predict-observe-explain structure of the worksheets. Then, student groups were split into sub-groups of maximum 12 students each and started their first of three experimental 90-minute workshops: the cloud chamber, electron tube or X-ray workshop. Three to four students work together as a team at one of the experimental setups. After lunch, students performed the two

remaining workshops. At the end of the day, students had the chance to ask last questions about their experiences at CERN or about the daily life of a scientist before they left S'Cool LAB and had the chance to visit CERN's Microcosm exhibition or the souvenir shop.

Data collection before, during and after the S'Cool LAB Day programme

To assess and model the educational effects of the hands-on session in S'Cool LAB, students were asked to fill out questionnaires before and after taking part in the S'Cool LAB Day programme. Due to time constraints, it was not possible to have students fill out longer questionnaires onsite on their S'Cool LAB Day. Instead, an online platform was used to administer most of the questionnaires.

During the pilot study (see chapter 6), students filled out the pre-visit questionnaires in the morning of the S'Cool LAB Day right before the start of their guided tour. The post-visit questionnaires were administered through the online platform. Due to technical constraints, however, one group filled out the post-visit questionnaire at the end of the day at CERN. During the main study (see chapter 7), students filled out online questionnaires administered via the online platform approximately 10 days before and 30 days after visiting CERN. In addition, directly after the guided tour, approximately half of the students filled out a short 10-minute questionnaire, which was used to assess the stability of their dispositional interest and self-beliefs.

Potential effects of the guided tour on the workshops in S'Cool LAB

Since the topics or the guided tour and the workshops overlap, it is possible that the experiences of the guided tour affect students experience in S'Cool LAB. For example, the projection mapping show in the synchrocyclotron does not explicitly mention the Lorentz force, the right-hand-rule or the differences between magnetic and electric fields. Nevertheless, students' experiences at this visit point might later affect how they interact with the electron tube experiment in S'Cool LAB. Also, students might recall facts about silicon detectors from their visit of the ATLAS visitor centre when working on the X-ray workshop, or remember basic principles of particle detection during the cloud chamber workshop.

4.5 S'Cool LAB workshops

In the framework of this thesis, different experimental workshops have been developed by the author and in collaboration with the S'Cool LAB team and visiting students. Three of these workshops formed the hands-on session in S'Cool LAB, which was the focus of the studies presented in chapters 6 and 7: the electron tube workshop, the cloud chamber

workshop, and the X-ray workshop. The following subsections summarise these three workshops including their key concepts. All student worksheets are included in the electronic appendix, see appendix 10.13.

4.5.1 Electron Tube - The Basics of Particle Acceleration

CERN is most famous for its particle accelerator complex, with the LHC (Large hadron Collider) being the most powerful accelerator in the world. Only one small section of the $27\ km$ circumference is used to increase the particles' energy using electric fields in radiofrequency cavities. Most of the ring in equipped with strong dipole magnets that bend the particle trajectory to return the particles back to the accelerating cavities. Although particle acceleration is probably the most important concept at CERN, students and the general public typically hold severe misconceptions about the acceleration of particles. In particular, many think that the magnets in the LHC are used to increase the particles' energy instead of merely deflecting the beams. This is also reflected in previously documented students' conceptions, see subsection 2.4.1.

With the help of an electron beam tube (or cathode ray tube), students study the behaviour of electrically charged particles in electric and magnetic fields. While students only operate a low-energy electron beam with an energy of $300\ eV$ per electron, they get to know the basics of particle acceleration in the LHC, in particular, particle acceleration through electric fields and deflection through magnetic fields. The LHC is a synchrotron. That means that the strength of the deflection magnets increases synchronously with the kinetic energy of the particles by increasing the electric current within large superconducting electromagnets. Similarly, this principle can be studied at a low-energy electron beam with electromagnets made of copper wire: the more energy the electron get from the acceleration, the stronger the magnetic field needs to be to keep the electrons on the same trajectory.

As a result, this workshop is structured as follows: First, a tutor explains the equipment and the associated safety hazards that come mainly from the operating voltage at the anode and the evacuated glass tube. They also explain how the electron source works, how electrons are accelerated by the anode, and how the Wehnelt cylinder helps to focus the electron beam.

Then, students connect the different components of the electron tube to the power supplies. Afterwards, they study the appearance of the electron beam when varying the acceleration voltage. Since the glass tube contains a small amount of a residual Neon gas, the beam looks orange due to the interactions of the accelerated electrons and the

electrons of the Neon atoms. Furthermore, a fluorescent coating at the end of the electron tube allows to visualise the electron beam spot in green. Next, students study the effect of a bar magnet on the electron beam and summarise their observation comparing it to right-hand-rule of the Lorentz force. After practising the applications of this rule to predict the deflection of electrons, students investigate the effect of magnets on electrons with different energies and learn about the basics of synchrotrons. As an optional task, students can connect a copper coil instead of a bar magnet and study its effect on the beam. A summary of the electron tube workshop for teachers, including the key concepts and required prior knowledge for this workshop, is included below.

Box 1. Summary of the electron tube workshop for teachers.
(available online at https://scool.web.cern.ch/content/preparation-electron-tube-workshops)

Summary of the electron tube workshop for teachers

What happens during this workshop?

Students learn how to make and control a particle beam and study how electrically charged particles behave in magnetic fields. Students learn about synchrotrons like the LHC.

Before taking part in this workshop, students should already know about:

- the properties of electrons and energy levels of electrons in atoms
- electric and magnetic fields, e.g., the shape of the magnetic field of a bar magnet and of a simple coil
- the deflection of electrically charged particles in magnetic fields due to the Lorentz force
- how to calculate the speed of electrons when they are accelerated by an electric field (optional)

Key concepts

1. Magnetic field lines: Magnetic field lines can be used to visualise magnetic fields. By convention they are defined going from the magnetic north pole to the magnetic south pole of a magnet.
2. Electric charges and electric fields: Electric field lines can be used to visualise electric fields. By convention they are defined going from + to -. Negatively charged particles are attracted by positively charged objects.
3. Lorentz force: Moving electrically charged particles can be deflected by magnetic fields. This is caused by the so-called Lorentz force.

Summary and link to CERN physics:

Electrically charged particles are deflected in magnetic fields. This is used in the LHC to make protons follow a circular path. To make particles faster, electric fields can be used. In the LHC, protons gain kinetic energy when passing through high electric fields (RF cavities). To keep the protons on a circular track, 1232 special superconducting electromagnets are installed. The high magnetic field of the dipole magnets causes the deflection of these electrically charged particles due to the Lorentz force.

4.5.2 Cloud Chamber - Build and Observe a Particle Detector

After the collision of high-energy protons in the LHC, gigantic particle detectors record the signatures of the new particles that have been created out of the collision energy. These particle detectors have an onion-like structure consisting of different layers of sub-detector systems, each specialised to measure a specific particle type or property. Many of these sub-detectors rely on ionisation processes in the detector material. One of the oldest particle detectors is the cloud chamber, which was the basis of the physics Nobel prizes in 1927 (Charles T. R. Wilson) and 1936 (Carl D. Anderson) and is also available to students in S'Cool LAB. Here, electrically charged particles leave a trail of ions when penetrating the super-cooled super-saturated isopropanol layer of a cloud chamber. These ions serve as condensation nuclei and trigger the condensation process, which leaves a trail of droplets that is easily observable even with the naked eye by shining a torchlight inside the cloud chamber. In S'Cool LAB, groups of three to four students assemble their own cloud chamber using dry ice blocks and isopropanol. First, they observe the random condensation processes above a cooled black aluminium plate. Once tracks of electrically charged particles are visible, students are asked to categorise the different types of tracks. With the help of a tutor, students assign the tracks to different types of particles from cosmic and terrestrial sources. In this way, basic concepts of particle detector are introduced without the need for digital data acquisition and analysis.

The research on students' conceptions of natural radiation including cosmic particles is quite limited, see subsection 2.4.2. Many students are not aware of cosmic sources of ionising radiation. Instead, many students hold the historic misconception that the soil is the only source of natural radiation. Consequently, they expect the amount of radiation to decrease when moving further away from the Earth, for example, when climbing a mountain.

In S'Cool LAB, cloud chamber workshops are not guided by worksheets, because the light conditions make it hard to read or write on paper and the amount of experimental manipulations is limited once the chamber is assembled. Instead, a tutor shows a brief presentation to explain the step-by-step instructions on how to build a cloud chamber highlighting the associated safety hazards due to dry ice and isopropanol. Afterwards, students assemble their own cloud chamber in teams of three to four. Next, they observe their chamber looking for first particle tracks appearing as thin white lines. Students are then asked to draw and categorise the observed tracks based on their different shapes. After disassembling the chambers, students and the tutor gather around a whiteboard to interactively discuss how the cloud chamber works and what types of particles can be

associated with which types of tracks. Tutors introduce students to the concepts of both terrestrial and cosmic sources of high-energy particles. In particular, Radon-222 isotopes in the air emit Helium nuclei (alpha radiation) when decaying into Polonium-218. Primary cosmic particles such as high-energy protons transform, for example, into (anti-) muons when traversing the atmosphere, which eventually transform into positrons or electrons together with corresponding (anti-) neutrinos. A summary of the construction manual and the physics background that is usually part of cloud chamber workshops can be found online (Woithe, 2016). A summary of the cloud chamber workshop for teachers, including the key concepts and required prior knowledge for this workshop, is included below.

Box 2. Summary of the cloud chamber workshop for teachers.
(available online at https://scool.web.cern.ch/content/preparation-cloud-chamber-workshops)

Summary of the cloud chamber workshop for teachers

What happens during this workshop?

Students learn about the principles of particle detection. They build a cloud chamber using the equipment provided, observe and categorize particle tracks, and discuss their observations with the rest of the group and a CERN scientist.

Before taking part in this workshop, students should already know about:

- different types of particles, such as proton, neutron, electron, photon, (muon, positron, neutrino) and their properties
- radioactivity and natural radiation (alpha, beta, gamma), ionisation
- the different states of matter, evaporation, condensation

Key concepts

1. Elementary particles: An elementary particle is a particle without substructure according to our current understanding. Examples of elementary particles are electrons, muons, and photons. We describe protons and neutrons as particle systems, because they are made of quarks. The alpha particle is a special system, because is consists of 2 protons and 2 neutrons – also known as a helium nucleus.
2. Particle properties: We distinguish particles by their properties, such as mass and electric charge.
3. Ionisation: Ionisation is the process by which the equal number of protons and electrons in a neutral atom changes: Either an electron leaves an atom after absorbing enough energy, or an electron joins an atom.

Summary and link to CERN physics:

Ionising particles are part of our everyday life although we can't see them. They can be made visible using relatively simple equipment. Cosmic particles colliding with the upper layers of the Earth's atmosphere produce showers of new particles, some of which reach ground level. Naturally-occurring radioactive elements in the air (Radon) emit alpha particles. The cloud chamber was used to discover the positron (Nobel Prize 1932) and the muon (Nobel Prize 1936). Today, cloud chambers are only used in education. Modern particle detectors at CERN (e.g., ATLAS) use similar principles of particle detection as cloud chambers (though different techniques) to explore the properties of particles.

4.5.3 X-rays - Medical Applications and Pixel Detectors

As an example of a modern particle detector technology, students use pixel detectors in S'Cool LAB (MX-10 detectors, www.particlecamera.com). These devices use a silicon chip (256 x 256 Pixel, total area $1.4 \times 1.4\ cm^2$) that was originally developed by the Medipix2 Collaboration at CERN, and is currently used in medical imaging. When high-energy particles fly through the semiconductor material, they create charge carriers (electrons and electron holes), which migrate to electrodes made of metal. This current is then amplified and evaluated. Due to the division of the detector area into pixels, pixel detectors can not only measure the energy deposited by particles but also their trajectory. In the LHC detectors, layers of pixel detectors serve as tracking detectors for the particles produced in the proton collision. Although CERN's main focus is fundamental research, technology developed for high-energy physics is also used in other areas, such as medical physics. Here, pixel detectors used for computer tomography measure the absorption rate of photons in different types of body tissue to reconstruct three-dimensional X-ray images of the human body.

Students typically compare X-rays with visible light, struggle to understand the different interactions that are relevant, and thus they confuse X-ray images and photographs (Clement & Fisseux, 1999). In particular, they often expect black paper to absorb X-ray photons more than white paper, because they do not know that only the electron density of the material determines the absorption probability of X-ray photons for the photoelectric effect. In contrast, many students believe that exposure to X-radiation turns stable atoms into radioactive atoms while relying on different lines of reasoning, see subsection 2.4.2 for a detailed overview of students' conceptions.

During the X-ray workshop in S'Cool LAB, a tutor first introduces students to the X-ray device and explains how to operate it safely. Then, students use a fluorescent screen to visualise the X-ray beam. Next, students identify a broken bone in a model of a human body and study the absorption of X-ray photons in different material to understand how the interactions of X-ray photons differ from photons from the visible spectrum, and how X-ray images are produced. After an introduction of the basic principles of pixel detectors and measurements of natural radiation, students design and perform a small experiment to test the posed hypothesis that food becomes radioactive when it is exposed to X-radiation. In the end of the workshop, students have the chance to X-ray different objects, such as small suitcases to simulate the X-ray scans of (potentially dangerous) luggage at the airport.

A summary of the X-ray workshop for teachers, including the key concepts and required prior knowledge for this workshop, is included below.

Box 3. Summary of the X-ray workshop for teachers.
(available online at https://scool.web.cern.ch/content/preparation-x-ray-workshops)

Summary of the X-ray workshop for teachers

What happens during this workshop?

Students learn more about high-energy photons and their interaction with matter. They understand how the X-ray machine works and observe the absorption of high-energy photons in different material. Students use pixel detectors to observe background radiation and to take radiographs.

Before taking part in this workshop, students should already know about:

- the properties of photons, especially the link between different photon energies and the electromagnetic spectrum
- how X-rays machines work, *Bremsstrahlung* (optional)
- semiconductor detectors (pixel detectors)
- energy levels of electrons in atoms (optional)

Key concepts

1. Energy spectrum: Different physics phenomena, such as visible light, X-rays, and radio waves can be described as electromagnetic waves with different wavelengths. In particle physics this corresponds to photons with different energies.
2. Absorption of photons: Electrons can absorb photons. If they absorb a photon, they take up their energy. If the absorbed energy is high enough, the electron will leave its atom (ionisation).

Summary and link to CERN physics:

High-energy photons produced by X-ray machines interact with matter (dominant effect: photo effect), as a result some of them are absorbed and we see a "shadow" behind strongly absorbing objects. Depending on their energy, photons interact differently with matter. For X-Ray photons, the electron density of the material correlates with the absorption rate. With the pixel detector one can measure particles all the time even without turning on the X-ray source – the so called background radiation (see cloud chambers).

Pixel detectors are used in the tracking systems of LHC experiments, such as ATLAS. They can precisely measure the track of an electrically charged particle. Within the Medipix collaboration at CERN, several applications of this technology have been studied, for example, in medical imaging.

5 Data preparation and analysis methods

A pilot study and a main study were conducted within this PhD project to analyse the effects of the hands-on session, which was part of the S'Cool LAB Day programme at CERN. Furthermore, a study on students' conceptions of radiation and radioactivity was conducted to analyse students' conceptual understanding in more detail. This chapter describes general considerations relevant for the data preparation and analysis of these studies. In particular, the following chapter describes standard methods used to analyse item properties according to classical test theory (section 5.1) and the steps taken to scan and to prepare the data for further analysis (section 5.2). Factor and cluster analysis are used to discover structure in the data and are described in section 5.3, while the methods used to analyse gains in students' educational outcomes are described in section 5.4. Sequential linear regression analysis is employed to model these educational outcomes based on students' novelty influence and experience factors (5.5), with multi-level linear modelling taking the clustering of data into account (5.6). Finally, section 5.7 describes the Bonferroni correction and its use across the different analysis methods to reduce the inflation of the alpha level, section 5.8 includes details about the sampling and section 5.9 summarise the software that is used.

5.1 Standard methods and parameters

Classical test theory provides several items and test criteria to evaluate the quality of a test while comparing them with recommended thresholds values. Standard criteria used in the framework of this project are summarised below.

5.1.1 Item mean, standard deviation and standard error

The (arithmetic) mean M_j characterises the average answer of all participants in the sample for item j, whereas the standard deviation SD_j quantifies the variability of the sample for item j around the sample mean M_j. Psychometric scales formed by six-level items (as used for affective variables, such as interest in this study) typically produce mean values between 1 and 6 and are thus often transformed linearly to the interval $[0, 1]$ to simplify their interpretation. In this case, the sample standard deviation is transformed in the same way. Indeed, Cohen, Cohen, Aiken, and West (1999) advocate for converting scores in this way to the percent of maximum possible score (POMP). When standardising variables by transforming them into z-scores, the mean is set to 0 and the standard deviation to 1. This standardisation is used, in particular, for linear regression analyses to simplify the interpretation and comparability of regression weights. Furthermore, the

standard error of the mean SEM_j for item j is estimated as the sample standard deviation SD_j divided by the square root of the respective sample size N_j.

5.1.2 Item easiness (or item difficulty)

The item easiness (often called 'item difficulty') P_j refers to the proportion of the sample who solved item j correctly and hence varies between 0 and 1. This parameter is usually called 'item difficulty' although many authors suggest that 'easiness' is the term representing the parameter more accurately (Ding & Beichner, 2009; France & Batchelder, 2015; Kaplan & Saccuzzo, 2009). Ding and Beichner (2009) recommend an item easiness for cognitive multiple-choice items in the interval $[0.3, 0.9]$. Items that are too easy (or too difficult) do not distinguish well between learners since almost all learners get these items right (or wrong). The lower bound for the item easiness values typically refers to the guessing probability, which depends on the number of answer options and hence needs to be adapted to 0.5 for Multiple-True-False items. The optimal item easiness usually lies around the halfway point between 100% and the guessing probability (Lord, 1952).

5.1.3 Item discrimination

Useful test items clearly distinguish between high and low performers and item discrimination indices characterise the discrimination power of an item j belonging to a test t. Kelley, Ebel, and Linacre (2002) compare the properties of different item discrimination indices for dichotomous data: the discrimination index D, which is computed according to the upper-lower 27% group method (Baker, 1965), the point-biserial correlation (which refers to the corrected item-to-total correlation r_{tj} in case of non-dichotomous data) and Rasch item discrimination. They summarize that all three discrimination indices give similar information but recommend using the point-biserial correlation or Rasch item discrimination. In the following, the **corrected item-to-total correlation r_{tj}** will be used to assess item discrimination both for dichotomous and non-dichotomous items. This parameter gives the Pearson product-moment correlation between item j and the total sum of all other items of a given test t. The value indicates how well each item correlates with the other items of a scale. Typically, correlations below 0.2 indicate that an item might not measure the same construct as the other items and thus should be discarded (Ding & Beichner, 2009; Kline, 2005)

5.1.4 Cronbach's alpha & Kuder-Richardson index

Cronbach's alpha α_C (Cronbach, 1951) is used as an estimate of the internal consistency of a scale or a unidimensional test. If items correlate high with each other, they probably

assess the same construct, which also leads to higher values for Cronbach's alpha. However, Cronbach's alpha needs to be used with caution, because it tends to increase with a higher number of items in a scale and it allows no conclusion about the dimensionality of a scale (Cortina, 1993).

As a rule of the thumb, Mallery and George (2003) suggest to interpret values of Cronbach's alpha as follows:

- $\alpha_C > 0.9$ Excellent
- $0.7 < \alpha_C < 0.9$ Good
- $0.6 < \alpha_C < 0.7$ Acceptable
- $0.5 < \alpha_C < 0.6$ Poor
- $\alpha_C < 0.5$ Unacceptable

However, Taber (2017) discusses that the interpretation of these values depends on the context of the test: For example, tests measuring conceptual understanding are usually associated with lower values for Cronbach's alpha, especially when they aim to test a broader range of aspects.

Although Cronbach's alpha is often misused in science education research, for example, as a measure for uni-dimensionality or over-interpreted as a measure for scale reliability (Taber, 2017), it can be a useful measure in particular for scale purification procedures, because it allows to assess the item-specific variance, the uniqueness of an item (Cortina, 1993). When calculating a series of Cronbach's alphas α_{tj} for a test t while omitting one of the test items j at the time, one can estimate if and how a scales' consistency improves when discarding items with a higher item-specific variance, i.e., items that differ from the other items of the scale.

5.2 Data screening and preparation

5.2.1 Missing data

In general, missing data were coded as such using the missing value code '999' in SPSS. Students were excluded from the respective analyses if they missed an item relevant to one of the scales of the analyses. However, it was assumed that the use of online questionnaires reduced the amount of missing data dramatically as students were forced to answer all items before submitting the questionnaire.

5.2.2 Normality

Many statistical tests such as the t-test assume that all continuous variables follow a normal distribution. Hence, analysing variables for normality, either using statistical or graphical methods is recommended before running inference tests (Tabachnick & Fidell, 2013). Whereas **kurtosis** describes the peakedness of a distribution, **skewness** describes the symmetry of the distribution. Both properties can be used to assess the normality of a given sample by analysing the deviation of both values from zero using the z-distribution, see equations in Table 7. In addition, both the Kolmogorov-Smirnov (KS) test and Shapiro-Wilk (SW) test can be used to compare a sample with a normal distribution. If the KS and SW statistics deviate significantly form zero, the assumption of a normal distribution may be violated. However, Tabachnick & Fidell (2013) recommend adding visual inspection, for example, of the shape of a distribution via expected normal probability plots when handling large sample sizes: Since the standard errors for skewness and kurtosis decrease with larger sample size, the null-hypothesis of normal distribution will often be rejected even if there are only minor deviations. For example, variance over- and underestimation effects due to positive or negative kurtosis disappear with more than 200 cases (Waternaux, 1976).

Table 7. Estimation of skewness and kurtosis values (Tabachnick & Fidell, 2013). Extreme z-values with $|z| > 3.29$ indicate a significant deviation from a normal distribution.

Skewness		Kurtosis	
Standard error for skewness s_S:	z-value for skewness value S:	Standard error for kurtosis s_k:	z-value for kurtosis value K[5]:
$s_S \approx \sqrt{\frac{6}{N}}$	$z = \frac{S - 0}{s_S}$	$s_k \approx \sqrt{\frac{24}{N}}$	$z = \frac{K - 0}{s_S}$

Data transformations can help to reduce effect of outliers or to improve normality, linearity and homoscedasticity. However, data transformations can make it much harder to interpret transformed variables, especially if the scales in which the variables are measured originally are well established. Furthermore, when analysing grouped data, transformations are not advisable (Tabachnick & Fidell, 2013). Another solution for variables with serious violations of the assumptions listed above can be dichotomisation into dummy variables.

[5] The kurtosis of any normal distribution is 3. However, SPSS reports an adjusted version of the kurtosis, a so-called excess-kurtosis, which is given by kurtosis minus 3. Therefore, the reported kurtosis value is compared with 0 instead of 3.

5.2.3 Coding errors, outliers and contaminants

Outliers are extreme observations that are inconsistent with the rest of the data set. As such, they can have significant and severe effects on several statistical methods, such as linear regressions (Pirker, 2009). The type of outlier that is easiest to find and to treat are usually coding errors that were entered into the data set by mistake. Comparing minima and maxima with the expected values for all items identifies this type of outliers.

A special type of outliers are contaminants, which are extreme cases (in this study: students) with several outlying variables indicating that they are not part of the targeted sample. To identify univariate outliers**,** composite variables (e.g., interest in physics) and single interval-scaled variables (e.g., experience with experiments) used for further analysis are transformed into standardised z-scores. Cases with $|z| > 3.29$ are analysed and in some cases handled as outliers by adjusting their extreme z-score to $|z| = 3.3$ (Tabachnick & Fidell, 2013). This z-score refers to 0.1% of the cases that are expected to lie outside this boundary conditions for normally distributed data. However, distributions deviating from a normal distribution are inspected carefully to avoid mistaking data points in heavy-tailed distributions with outliers (Tabachnick & Fidell, 2013).

Multivariate techniques such as multiple linear regressions require analysis of **multivariate outliers** in addition to univariate outliers, for example, by visually inspecting scatter plots, by analysing the **Mahalanobis distance** (Mahalanobis, 1936), or by calculating Cook's distance, leverage, the Minimum Covariance Determinant or similar measures (Leys, Klein, Dominicy, & Ley, 2018). Analysing outliers in multi-level data is even more complex because, technically, outliers need to be analysed on different levels. For example, level-1 variables need to be analysed within the respective level-2 cluster (not with regard to the sample mean), which is seldom feasible for large cluster numbers. Furthermore, treating outliers on one level might affect the status variables on other levels (Langford & Lewis, 1998). Therefore, outliers were only analysed with respect to the sample mean and a simplified strategy for outliers, loosely based on Tabachnick and Fidell (2013), is applied:

1. Compare minima and maxima of all numeric items with expected range to avoid coding errors
2. Check for reasonable means and standard deviation
3. Standardise univariate data (composite and single interval-scaled variables) into z-scores
4. Identify all extreme values with respect to the sample mean using the critical z-score $|z| > 3.29$
5. For outliers: adjust values to less extreme values

6. For extreme values in heavy-tailed distributions: report the extreme value without changing it, but list group identify of the case
7. Check for group-level effects by checking for clustering of extreme values
8. Calculate the Mahalanobis distance for multivariate analysis methods and exclude multivariate outliers from the respective analyses
9. For regression analyses: analyse the standardised regression residuals for extreme cases with $|z| > 3.29$
10. Compare results of multivariate analyses with and without extreme cases to estimate their leverage
11. Analyse students with outliers in several variables to identify contaminants not belonging to the modelled sample

5.3 Discovering structures in S'Cool LAB's educational data

Factor and cluster analysis are used to identify structures in the data gathered within the framework of S'Cool LAB workshops. While factor analysis is used to discover structures in a number of items in questionnaires, cluster analysis is used to discover structures in the student sample.

5.3.1 Factor analysis

Psychometric measurements often rely on several items measuring different aspects of the multi-facet latent concepts, which cannot be observed directly, such as 'interest in physics' or 'physics-related self-beliefs'. Here it is assumed that the extent to which an individual is, for example, interested in physics causes the individuals' replies to the developed test items.

Here, factor analysis is a widely used statistical method analysing the variability of several observed and correlated variables and aims at describing this variability through a lower number of unobserved, independent, underlying variables, so-called factors (Tabachnick & Fidell, 2013). These factors are assumed to represent underlying concepts creating certain patterns of correlation among the variables. Observed variables are then modelled using linear combinations of these factors with parameters called factor loadings and error terms. Moreover, the squared factor loadings represent the percentage of variance in the observed variable, which is modelled by the corresponding factor.

Thus, factor analysis can help to develop consistent measurement instruments for abstract concepts by evaluating the properties of single items with respect to their factor fit.

Furthermore, factor analysis helps to assess the validity of an instrument by testing whether the instrument is indeed structured by the hypothesized factors.

Two major types of factor analysis are distinguished based on whether the number of factors is defined beforehand or not: **Confirmatory factors analysis** assumes a certain number of underlying factors based on the underlying theory or prior results, which is then tested on a data set. The analysis will produce factor loadings of observed variables on this fixed number of latent factors. If no or little theoretical background can be used to predict the relationship between different variables, **exploratory factor analysis** can be used to explore the factor structure without assumptions about the number of factors (Tabachnick & Fidell, 2013).

For exploratory factor analysis, the number of relevant factors can be estimated by considering only extracted factors with eigenvalues greater than 1. Furthermore, Cattell (1966) suggests analysing the scree plot, a plot of the eigenvalues of all extracted factors in decreasing order. When following the distribution of eigenvalues from the right (small eigenvalues) to the left (large eigenvalues), a 'kink' indicates the last factor that is considered to be part of the scree at the base of the 'eigenvalue cliff'. Only factors with eigenvalues lying left of this kink are considered substantial. Although the evaluation of the kink is somewhat subjective, Fabrigar, Wegener, MacCallum, and Strahan (1999) consider the scree-test suitable if there are strong common factors in the data. Furthermore, Osborne, Costello, and Kellow (2008) suggest to compare factor solutions for different numbers of extracted factors and chose the solution best fitting the data. They describe this solution as the 'cleanest' factor structure characterised by the smallest number of cross-loadings. Here, all factors should have at least three items with item loadings higher than 0.32.

Factor extraction, rotation and requirements

The two most often used factor extraction techniques are principal axis factoring and principal components analysis (Tabachnick & Fidell, 2013). Although both techniques typically lead to only small differences in their outcomes, in particular, in the magnitude of the obtained factor loadings (Russell, 2002), their theoretical and mathematical backgrounds differ. Here, Tabachnick and Fidell (2013) describe major theoretical and mathematical differences between both techniques in detail. In short, **principal axis factoring** assumes an underlying theory causing the correlation between certain variables and analyses only the covariance between variable. This approach is recommended when analysing variables that have been designed based on hypothetical underlying constructs. In contrast, **principal components analysis** identifies patterns of correlated variables by

analysing all the variance of the variables. Although principal components analysis seems to be more popular, principal axis factoring is recommended by Russel (2002) because of its more generalizable method of factor extraction compared to principal components analysis.

While principal axis factoring attempts to maximise the modelled variance, this technique can lead to several items loading on more than one factor. Here, rotation methods help to make the output easier to interpret. The **Varimax rotation** is the most common orthogonal rotation method that maximises the variance of factor loadings by increasing high factor loading and decreasing low factor loadings while keeping the factors uncorrelated with one another.

Typically, factor analyses require a relatively large sample size of at least 100 cases with at least three items per factor to produce a stable factor solution (Russell, 2002; Tabachnick & Fidell, 2013). Osborne et al. (2008) analysed 303 case-to-item ratios in exploratory factor analyses and find that 41% of the studies report a ratio $\leq 5:1$. To decide whether data is suitable for factor analysis, the **Kaiser-Meyer-Olkin (KMO)** measure of sampling adequacy indicates the proportion of variance in the variables that could be attributed to underlying factors by comparing partial correlations between pairs of variables with the original correlations. KMO values > 0.8 are desirable, whereas KMO values < 0.5 characterise data that is not suitable for factor analysis (Cerny & Kaiser, 1977).

Factor analysis on binary data

While standard factor analysis techniques are based on Pearson correlation matrices, Multiple-True-False (MTF) items produce dichotomous data and hence pose a special case to factor analysis because binary data violates the assumption of a continuous distribution. Although it is technically possible to proceed with standard factor analysis procedures using Pearson correlation in this case, it is inappropriate, and the interpretation of factors typically suffers from several problems (Galbraith, Moustaki, Bartholomew, & Steele, 2008). In particular, factor loadings are usually downwardly biased because the Pearson correlation coefficients on dichotomous data seldom span the range from -1 to $+1$ (StataCorp, 2019). Furthermore, factor analysis might produce additional factors with no conceptual interpretation, so-called difficulty or skew factors (Lorenzo-Seva & Ferrando, 2012).

Here, the **tetrachoric correlation** (Pearson, 1900) can serve as a better-suited measure of association for factor analysis. It assumes that the association between two dichotomous variables is based on unobserved underlying continuous variables – as it is

the case for Multiple-True-False items testing students' conceptions. Therefore, the tetrachoric correlation is a widely used estimate of the product-momentum correlation of these underlying continuous variables (Lorenzo-Seva & Ferrando, 2012). The tetrachoric correlation assumes and estimates certain thresholds for the underlying continuous variables, which dichotomise the variable and generate the observed data. However, this estimate is less stable than the Pearson correlation and should be used with caution only on larger samples of at least 200 to 300 observations (Guilford, 1950). Galbraith et al. (2008) report similar results when comparing factor analysis on a tetrachoric correlation matrix and a logit latent variable model and argue that these two procedures are indeed equivalent for binary data.

5.3.2 k-means cluster analysis

While factor analysis detects how items group together, clustering can be used to find out how data points (i.e., individual learners) group together. Clustering describes the process of partitioning a set of data objects into groups, so-called clusters, so that all objects within one cluster are similar to each other but differ from objects belonging to different clusters (Han, Kamber, & Tung, 2001).

Cluster analysis is a widely used data mining approach. For example, it identifies different types of customers with similar characteristics to later target specific clusters of customers with adapted business strategies (Han et al., 2001). Furthermore, clustering can help to improve instruction by discovering structure in educational data (Baker & Inventado, 2014).

Different methods (and algorithms) are available to identify groups within data sets, such as hierarchical methods, partitioning methods or density-based methods (Han, Kamber, & Pei, 2011). The **k-means algorithm** is an example of a partitioning method. It creates k initial clusters within a data set of M N-dimensional data objects and sorts these data objects into the nearest cluster based on, for example, the Euclidian distance to the initial cluster centre. However, the researcher needs to define the number of clusters k as well as the initial set of cluster centres in advance (or set cluster centres randomly). An *'iterative relocation technique'* then determines the new cluster centres based on the mean within each cluster. Afterwards, it moves data objects from one cluster to another cluster to optimise the clustering by minimising the sum of squares within one cluster. Convergence is reached when the cluster centres stop changing between iterations (Kaufman & Rousseeuw, 2009). However, it is difficult to evaluate the quality of clustering results because the quality parameters are part of the optimisation problem. Hence, interpreting clusters and their usefulness relies on human evaluation (Estivill-Castro, 2002).

5.4 Analysis of gains & differences

Educational interventions usually try to optimise students' educational outcomes. When student variables are measured before and after an intervention, t-test can be used to compare the variable means. While t-tests evaluate the statistical significance of group differences (or: how surprising is the effect?), effect sizes quantify the size of the differences (or: how wrong would it be to reject the alternative hypothesis H1?).

5.4.1 Comparing group means with t-tests

Several statistical hypothesis tests can be used to assess the statistical significance of the difference between two sample means. The *t*-tests are based on a test statistic following the Student's t-distribution under the null hypothesis that the population means are equal. The Student's t-distribution is a continuous probability distribution used when estimating the mean of a normally distributed population in cases with relatively small sample sizes and an unknown standard deviation (Student, 1908).

The t-tests require a continuous normally distributed variable, two different groups and randomly selected samples of a population. Education research seldom meets all these criteria. However, t-tests are quite robust against violations of these criteria if the sample size is big enough and relatively similar for the two groups (Muijs, 2004). Indeed, according to the central limit theorem, the normalised sum of independent random variables tends toward a normal distribution if the sample size is large enough even if the original variables are not normally distributed.

If the sample sizes differ between two groups, **Levene's test** (Levene, 1961) is calculated first to check the homogeneity assumption that the population variances of the dependent variable are equal for all groups. In particular, this test tests the null hypothesis that the groups have equal population variances. If the null hypothesis is rejected, the degrees of freedom df of the t-test are adjusted accordingly.

5.4.2 Effect size Cohen's d

While statistical tests like the t-test assess the statistical significance between two groups, effect sizes are a way to calculate and compare the size of the differences. Specifically, when discussing students' learning gains, a normalised gain is often reported instead of the effect size. However, Hake's gain is biased against low pre-test scores, whereas Cohen's d_C is better-suited for pre-post comparison because it takes the correlation between these two measures into account (Nissen, Talbot, Thompson, & Van Dusen, 2018).

If $\bar{x}_1$ and $\bar{x}_2$ are the means of two different samples and $\bar{s}_1^2$ and $\bar{s}_2^2$ are estimated variances of each sample, then Cohen's d_C is calculated as follows (Cohen, 1988):

$$d_C = \frac{Mean\ for\ group\ 1 - Mean\ for\ group\ 2}{Pooled\ standard\ deviation} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_p} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{2}}} \quad (1)$$

Cohen (1988) defines the effect size of the group differences in behavioural science according to the combined area covered by two equal-sized equally varying normal distributed populations, which is not overlapping:

- Small effect $d_C = 0.2 - 0.4$, 15% not overlapping
- Medium effect $d_C = 0.5 - 0.7$, 33% not overlapping
- Large effect $d_C \geq 0.8$, 47% not overlapping

Hattie (2008) sets his benchmark for comparing more than 800 meta-analyses related to student achievement at $d_C = 0.4$, the average effect size. He considers effect size above this hinge-point as desirable. Furthermore, effect size are usually reported together with the respective confidence intervals when using statistical significance testing (Cohen, 1988; Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999).

Repeated measures design

When analysing effects of single-group pre-post-test designs, the compared variables do not originate from independent (or separate) groups. Therefore, Morris and DeShon (2002) suggest a more appropriate estimate that takes the correlation ρ between the pre- and post-test measures into account. Here, the pooled standard deviation for repeated measures σ_{RM} includes the correlation ρ (equation 2), which is then used to obtain the corresponding Cohen's d_{RM} (equation 3). If the correlation between pre- and post-test variables is larger than 0.5, the effect size d_{RM} obtained in this way will be larger than the independent group effect size d_C.

$$\sigma_{RM} = \sigma_p \cdot \sqrt{2 \cdot (1 - \rho)} \quad (2)$$

$$d_{RM} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{RM}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(1 - \rho)(s_1^2 + s_2^2)}} \quad (3)$$

5.5 Multiple linear regression analysis

Linear regression analysis aims to model relationships between one or more independent variables (called predictors) and a dependent variable using linear predictor functions. The model parameters of these predictor functions are estimated from the data. In science education research, the goal is often to model variability in the dependent outcome variable that can be associated with variability in the predictor variables. However, linear regressions allow no conclusions about causal relationships[6].

The parameters of the linear regression models can be used to compare the predictive power of several predictor variables. Given a data set of answers by n students to p independent variables $\{x_{i1}, \dots, x_{ip}; i = 1, \dots, n\}$ and the observed dependent variable $\{y_i; i = 1, \dots, n\}$, and assuming a linear relationship between the dependent variable, the predictors, the model takes the form

$$y_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} + \varepsilon_i \quad i = 1, \dots, n \tag{4}$$

The parameter b_0 can be interpreted as intercept term, whereas $b_1, \dots, b_p$ are called regression coefficients and ε_i represents the error term. The goal of regression analysis is to obtain a set of regression coefficients while minimising the (sum of the squared) residuals.

The values of the regression coefficients b_i depend on the scale on which the predictors are measured, which complicates comparing the effects of different predictors. Therefore, **standardised regression coefficients** are obtained by standardising (z-transforming) all variables to a mean of zero and a standard deviation of 1. Also, standardised regression coefficients β_i can be calculated as follows from regression coefficients

$$\beta_i = \frac{b_i \sigma_{x_i}}{\sigma_y} \tag{5}$$

where σ_{x_i} and σ_y refer to the standard deviations of the independent variable x_i and the dependent variable y. A value of $\beta = 2$ would indicate that increasing the independent variable by one standard deviation will increase the dependent variable by two standard deviations.

[6] The terms ‘explained variance’ and ‘modelled variance’ are used somewhat interchangeably, although regressions cannot be used to establish causal relationships. Hence, to support linguistic accuracy, the word ‘model’ is used in the following when describing results of regression analyses as suggested, for example, by Snijders (1994).

5.5.1 Sequential linear regression

In sequential (often called hierarchical) regression, the independent variables are inserted one by one into the model. In particular, when handling correlated predictor variables, it is important to carefully decide on the way they enter a linear model because variables inserted earlier might account for the overlapping variance between the predictor variables. Thus, their importance as measured by regression coefficients, might be overestimated. Consequently, the order of predictors is defined by the researcher, for example, by considering their assumed time of causation: 'older' and more stable variables, such as gender or dispositions enter the model before 'newer' and less stable variables, such as personality states. Also, the theoretical importance of predictors can inform the order giving priority to theoretically important predictors. Last but not least, the research question can affect this decision. For example, it might be interesting to understand the effect of a variable after eliminating the effects of other variables (Tabachnick & Fidell, 2013). In this way questions such as "*Which relationship exists between situational interest and gender if the effects of dispositional interest are eliminated?*" can be studied. In addition, it is possible to compare the ability of groups of independent variables to model a dependent variable. For example, researchers can compare easily obtained variables such as 'Physics grade in last grade report' with more difficult and time-consuming measures such as 'test score of a concept test', or compare the predictive power of novelty influence versus novelty experience variables on educational outcomes.

A series of regression models can also help to identify mediator variables. Baron and Kenny (1986) describe the following conditions that must be met to consider a mediator variable between an independent variable IV and a dependent variable DV (see also Figure 13):

- Regression model 1 (IV on mediator, **only path a**): the independent variable IV significantly predicts the mediator variable
- Regression model 2 (IV on DV without mediator, **only path c**): the independent variable IV significantly predicts the dependent variable DV
- Regression model 3: (IV & mediator on DV, **paths a, b & c**): the mediator variable significantly predicts the DV (**path b**), while the IV predicts the DV significantly less compared to model 2. Path coefficient c decreases, or even goes to zero for a perfect mediator.

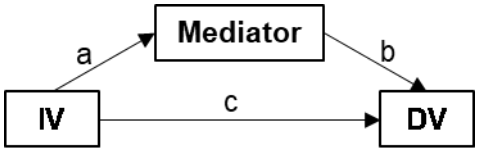


Figure 13. Path model for independent variable IV, dependent variable DV and a mediator variable.

However, Baron and Kenny (1986) emphasise that these conditions only point to potential potent mediators – they are not necessary and sufficient conditions to confirm an effect. Moreover, they point out the advantages of structural equation modelling for this type of models.

Contrary to the theory-guided strategy in sequential linear regression, the **statistical regression** is guided by purely statistical criteria such as correlation with the dependent variable or the additional modelled variance. It can be used to identify a subset of variables that is useful in modelling the dependent variable while eliminating other variables, which do not contribute significantly to the model. Although this exploratory technique can be helpful for model-building, Tabachnick and Fidell (2007) recommend caution, the use of large sample sizes, or cross-validation with a second sample before interpreting statistical regressions.

5.5.2 Assumptions of linear regression analysis

Regression analyses pose a special case for data screening. They require continuous variables or dummy coded categorical variables and are based on several assumptions regarding the data quality. The following steps to test these assumptions are based on (Tabachnick & Fidell, 2013).

The ratio of cases to independent variables and the absence of outliers

As a rule of thumb for the desired ratio of cases N to the number of independent variables m, Tabachnick and Fidell (2013) report equation (6)

$$N \geq 50 + 8m \tag{6}$$

when testing multiple correlation and equation (7)

$$N \geq 104 + m \tag{7}$$

when testing individual predictors (or the maximum of these two rules if testing both multiple correlations and individual predictors). These rules enhance the chances of a meaningful result when handling independent variables with medium-sized correlations to the dependent variable while assuming typical values for type I errors $(\alpha = 0.05)$ and type II errors $(\beta = 0.2)$. However, a higher case number is needed if the distribution of the

dependent variable is skewed, only a small expected effect size is expected, or less reliable variables are included. Moreover, to reduce the number of independent variables, it is possible to combine several highly correlated independent variables to form a new composite variable.

Outliers can affect the solutions of regression analyses as well as the precision of the estimated regression coefficients. Hence, all dependent and independent variables need to be analysed for univariate outliers, for example, by comparing their z-values to a value that is considered extreme, such as ± 3.29 (see also subsection 5.2.3). Furthermore, multivariate outliers can be identified, for example, by calculating and analysing the Mahalanobis distance (Mahalanobis, 1936). Consequently, regression analyses can be compared between the original data set and a reduced data set in which outliers are deleted.

Normality, linearity, homoscedasticity, and independence of residuals

Both linearity and homoscedasticity are important concepts when planning regression analyses. The assumption of **linearity** describes the need for a linear relationship between two variables. **Homoscedasticity** describes the homogeneity of the variance of random variables. To evaluate normality, linearity and homoscedasticity, Tabachnick and Fidell (2013) suggest inspecting the residual scatterplot, which displays the standardised regression residuals over the standardised predicted regression values. If the data meets all assumptions, the residuals will be distributed rectangular (without curvature) with a concentration of data points along the centre, a horizontal line at zero, and a normal distribution of residuals when going symmetrically away from this line. In this case, it is not necessary to scan single variables for normality (Tabachnick & Fidell, 2013).

Although a violation of linearity and homoscedasticity of the residuals will weaken the analysis, the results will still be valid. In particular, the goodness of fit may be overestimated (Tabachnick & Fidell, 2013). Also, the estimation of predictors is robust against violation of the assumption of normal distribution since this violation does not affect the regression model itself, but only the p-values for significance testing for small sample sizes. If the sample size is big enough, the estimation of standard errors and significance tests are not biased. Furthermore, the central limit theorem ensures that the disturbance term will be approximately normally distributed if the sample size is large enough (Eid, Gollwitzer, & Schmitt, 2017).

To test the residuals from least squares regressions for the **independence of observation** versus autocorrelations the Durbin Watson (DW) statistic is used. If the DW value is 2, no

autocorrelation is detected in the data set. A DW value between zero and 2 indicates positive autocorrelation whereas a DW value between 2 and 4 indicates negative autocorrelation. As a rule of a thumb, DW values between 1.5 and 2.5 are desirable (Field, 2009).

Absence of outliers in the solution

Extreme residuals indicate cases that are poorly fit by the regression model and are considered 'outliers in the solution'. Extremeness depends on the sample size, however, Tabachnick and Fidell (2013) suggest the criterion $p = 0.001$ translating into a z-value of ± 3.29 for sample sizes smaller than 1000. Extreme cases should be analysed in detail to understand limitations of the linear model and can be removed from the solution to increase multiple correlations.

Multicollinearity

A linear regression model will fit best when every independent variable is strongly correlated to the respective dependent variable but uncorrelated to other independent variables. In many cases, it is desirable to identify the lowest number of independent variables, each of which modelling a significant proportion of variance of the dependent variable. Especially in real-world problems, several predictors can be correlated and often it is not possible to use only orthogonal variables or research designs. Nonetheless, due to the flexibility of regression techniques, the linear model can still be applied. However, in this case the interpretation of regression coefficients becomes more complicated. In particular, the values for regression coefficients will decrease significantly when adding a new predictor that is highly correlated to a previous predictor variable. This effect can also be reflected by very small but still significant regression coefficients.

Tabachnick and Fidell (2013) recommend evaluating correlations between predictors, and consider omitting individual variables if their correlation is 0.70 or more. Furthermore, SPSS allows evaluating the **'tolerance' statistic** as part of its collinearity statistics. The tolerance statistic indicates the proportion of variance an independent variable shares with the other independent variables in the model:

$$Tol_j = 1 - R_j^2 \tag{8}$$

In equation 8, R_j^2 is the coefficient of determination of a regression model with the predictor variable j as dependent variable and all other predictor variables as independent variables. A tolerance value of 0 would indicate perfect multicollinearity, with the predictor j as a perfect linear combination of the other predictor variables. In contrast, a tolerance value of

1 would indicate that the predictor j is exactly orthogonal to the other predictors. As a rule of thumb, values smaller than 0.10 can indicate a multicollinearity problem (O'brien, 2007).

5.5.3 Model fit parameter

An important statistic derived from regression analysis is the Pearson correlation coefficient between the modelled and observed values for the dependent variable $r_{Y'Y}$, and its squared counterpart, the coefficient of multiple determination R^2. The latter describes the proportion of variance of the dependent variable that is accounted for by the independent variables. Adding independent variables to a linear model will never decrease but potentially increase the value of R^2. Therefore, an adjusted value of R^2 is often reported in addition, which penalises a high number of predictors m in a model or a small sample size n (Wherry, 1931):

$$adjusted\ R^2 = 1 - (1 - R^2)\frac{n-1}{n-m-1} \tag{9}$$

The adjusted R^2 gives an indication of the model fit. Muijs (2004) suggests the following interpretation:

- adjusted $R^2 < 0.1$: poor fit
- adjusted R^2 between $0.1 - 0.3$: modest fit
- adjusted R^2 between $0.3 - 0.5$: moderate fit
- adjusted $R^2 > 0.5$: strong fit

5.6 Multilevel linear modelling

Multilevel linear modelling is used to analyse data from participants that are not independent from each other but clustered in certain groups or organised on other higher levels (Hox, Moerbeek, & Van de Schoot, 2017; Tabachnick & Fidell, 2013). For example, students are typically part of a class, which is taught by a specific teacher in a certain school, which is located in a country. Furthermore, different predictors of the dependent variable may belong to different levels of the model. For example, prior knowledge can be considered a student-level predictor, while class size, the motivation of the teacher, funding of the school, or the curriculum of the respective country can be considered higher-level predictors, all of which may contribute to a student's interest in physics. Although many studies in physics education research analyse clustered data (e.g., by using whole classes of students from different schools), many researchers use only single-level linear modelling ignoring all nesting in the datasets (Van Dusen & Nissen, 2019). However, this type of

analysis assumes independence of error, which is often violated in student groups sharing the same teacher or the same learning experiences in a class. In short, students in one classroom tend to be more similar to each other than students selected from different classrooms or schools. Therefore, analysing multi-level data in a single-level model might lead to severe statistical challenges, such as an inflated type I error due to overestimation of the effective degrees of freedom. In addition, this might cause critical interpretational problems when drawing conclusions on one level based on data from another level. Here, the most prominent examples are the ecological fallacy and the atomistic fallacy (Diez-Roux, 1998). As Maas and Hox (2004) summarise: *"The consequences of using uni-level analysis methods on multi-level data are well known: the parameter estimates are unbiased but inefficient, and the standard errors are negatively biased, which results in spuriously 'significant' effects".*

Multi-level models, on the contrary, not only take individual differences into account, but also account for the differences between groups of individuals, which makes these models more suitable for many studies in school contexts. Indeed, instead of assuming randomly selected individuals, multi-level models assume that groups have been randomly selected from a population of groups. Multilevel models account for the nested data structure by including random effects that model how relations between predictors and outcomes differ between different groups of individuals. In addition to random effects, fixed effects model relationships between predictors and outcomes variables that are independent of the different groups. There are various options to account for multi-level data in linear models of varying complexity.

For example, **random intercept models** allow the means of a dependent variable to vary between higher-level units. Indeed, students taking part in S'Cool LAB workshops on one day might enjoy their experience more than students taking part in the same workshops on another day, because their teacher was very enthusiastic about their upcoming visit and thus motivated and prepared her students for the past 6 months, because the weather was better, or because the students slept better in their accommodation. Table 8 describes the standard equations relevant for random intercept models.

To determine whether the intercept varies between groups, multilevel analysis fits the model described in equation 10 to the data and compares this model with a model that does not include the group effect u_j on the intercept. A likelihood-ratio test then assesses the goodness of fit to determine the better model. Because maximum likelihood (ML) estimators are biased downwards for estimation of variance components, the restricted maximum likelihood (REML) approach is often applied. REML is a special form of

maximum likelihood estimating, which takes the number of fixed effects that need to be estimated into account (Harville, 1977).

Table 8. Standard equations for a random intercept model.

Equation:	$\boldsymbol{y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + \varepsilon_{ij}}$	(10)
Fixed part:	$\beta_0 + \beta_1 x_{ij}$	(11)
Random part:	$u_j + \varepsilon_{ij}$	(12)

Symbol	Meaning
y_{ij}	Dependent variable score for a case at level 1 (i.e., one student), i indicates the individual within a group, j indicates the group
β_0	Average intercept over all groups
β_1	Slope for the relationship between the level-1 predictor and the DV
x_{ij}	Score of level-1 predictor for a case at level 1
u_j	Deviation of the intercept from the mean intercept for group j
ε_{ij}	Random error of prediction for level-1 equation

Random slope models allow the regression coefficient to differ between higher-level units. In addition, variables related to the same or to different levels might be allowed to interact in **cross-level interaction models**. The underlying theory and the research questions typically determine whether random intercept models are sufficient or whether to include random slopes or even cross-level interactions. However, multilevel models include a much higher number of parameters that need to be estimated than single-level models. Therefore, larger sample sizes are required, and certain model types might not converge if the data structure is not favourable. However, some of the cluster sizes can be very small, as long as other clusters contain more individuals (Tabachnick & Fidell, 2013). Hox (1998) provides a rule of thumb when to consider which type of model depending on the number of clusters and their average cluster size, see Table 9.

Table 9. Recommendation for number of level 1 individuals per level 2 group (Hox, 1998).

Research interest	avg. cluster size	# clusters
Fixed parameters	30	30
Cross-level interactions	50	20
Random part, variance & covariance components	100	10

5.6.1 Assumptions of multilevel linear modelling and evaluation parameters

The assumptions for multilevel linear models are similar to ordinary linear models (linearity, homoscedasticity, normal distribution of residuals, see subsection 5.5.2). However, technically, the screening for uni- and multivariate outliers needs to be repeated for every level and every cluster, which is difficult for large cluster numbers (Tabachnick & Fidell,

2013). Besides, it becomes difficult to interpret the significance of predictors, especially in random slope models, because the variance of regression coefficients is part of the research question. If regression coefficients vary significantly over clusters, they might be very important for the model. However, their p-value might be well above typical thresholds. Moreover, cross-level interaction might increase multicollinearity issues. Therefore, the number of predictors should be chosen carefully and predictors should be entered stepwise in order of importance and left out if they do not improve the model (Raudenbush & Bryk, 2002). Finally, multi-level models are more complex than ordinary linear regression models and require the estimation of more parameters. Hence, larger sample sizes are required.

As for sequential linear models, the estimation of fixed effects in multilevel modelling is robust against violation of normal distribution. However, the standard errors for the regression coefficients are slightly biased downwards, in particular, if the number of level-2 units (e.g., groups) is less than 50 (Maas & Hox, 2004).

Contrary to ordinary linear models, the violation of the assumption of independence of errors is the justification of multilevel modelling. Indeed, the **variance partition coefficient**[7] (VPC) gives a direct measure of the dependence of errors. As such, it is also used to decide whether or not to account for higher-levels when analysing clustered data.

In a simple two-level model, it gives the ratio between variance attributed to the group level σ_u^2 and the total variance of a dependent variable, which is the sum of the group level (level-2) variance σ_u^2 and the individual level (level-1) variance σ_e^2.

$$VPC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \tag{13}$$

High VPC values indicate a severe violation of the independence of errors assumption, which translates into inflation of Type I errors (false positive conclusions) when proceeding with single-level models. As a rule of the thumb, Mehmetoglu and Jakobsen (2016) recommend multi-level modelling if the variance partition coefficient exceeds 5%.

An additional factor, the **variance inflation factor** (VIF, also called design effect) can be calculated to estimate the effect of the variance partitioning when considering clustering of data with an average cluster size c (Hox, 1998; Ukoumunne, Gulliford, & Chinn, 2002).

[7] This parameter is often called intra-class correlation. However, in this context it is not a correlation but in the order of variance. Therefore, the term variance partition coefficient is used instead as suggested, for example, by (Goldstein, Browne, & Rasbash, 2002).

$$VIF = 1 + (c - 1) \cdot VPC \tag{14}$$

The variance inflation factor describes how much the sample size would be overestimated when using a single-level model on clustered data. An overestimated sample size will, in turn, lead to an underestimation of the standard errors, which will inflate alpha levels when running significance tests. Therefore, the VIF allows computing the effective sample sizes.

5.6.2 Model fit criteria

Akaike's information criterion AIC (Akaike, 1974) and the **Bayesian information criterion BIC** (Schwarz, 1978) are often used to compare the relative quality of different statistical models. Both so-called 'penalised likelihood criteria' are based on the likelihood function and help to select a model out of a set of models with different dimensions that are all supposed to fit a given set of observations. Contrary to the likelihood, AIC and BIC include a penalty term for the number of parameters of a model, which prevents overfitting. Although AIC and BIC are closely related, they rely on different assumptions and show different performance depending on the data set and the model selection. For example, using AIC involves some probability of selecting models, which are too complex. Therefore, it is recommended to use them together by trying to find models that are favoured by both criteria (Kuha, 2004). Hence, when comparing different models, the model with the lowest AIC and BIC is preferred.

Because individual values for AIC and BIC are not interpretable as they contain several constants and depend on the sample size, it is recommended to rescale them before drawing conclusions to

$$\Delta AIC_i = AIC_i - AIC_{min} \quad \text{and} \quad \Delta BIC_i = BIC_i - BIC_{min} \tag{15}$$

This transformation identifies the best model when $\Delta BIC_i = 0$ or $\Delta AIC_i = 0$, while other models have positive values. Burnham and Anderson (2004) and Kass and Raftery (1995) suggest how to interpret BIC and AIC differences between different models as follows:

- ΔBIC/AIC of 0 to 2: Not worth more than a bare mention, original model supported
- ΔBIC/AIC of 2 to 6: Positive evidence against higher BIC/AIC model
- ΔBIC/AIC of 6 to 10: Strong evidence against higher BIC/AIC model
- ΔBIC/AIC of >10: Very strong evidence against higher BIC/AIC model

Although the AIC and BIC criteria are useful to compare different models, they do not provide absolute values for the goodness of the model fit. In ordinary linear regression analysis, R^2 and the adjusted R^2 describe the amount of variance of the dependent

variable that is accounted for by the predictors. However, estimating a coefficient of determination in multilevel modelling is challenging (Nakagawa, Johnson, & Schielzeth, 2017). Nevertheless, Van Dusen and Nissen (2019) and Snijders and Bosker (1994) suggest using a version of R^2 to select the best model which treats proportional reductions in the estimated variance components as **analogue R^2**:

$$analogue\ R^2 = 1 - \frac{sum\ of\ estimated\ variance\ components\ after\ inserting\ IVs}{total\ variance\ of\ DV} \quad (16)$$

However, there are a few problems with this approach: as this coefficient does not penalise the number of predictors, the number of predictors needs to be monitored manually. Van Dusen and Nissen (2019) suggest including only variables that explain at least 1% of the level-1 or level-2 variance. Furthermore, adding predictors to the model might decrease some of the variance components and might even cause negative values for R^2 (Snijders & Bosker, 1994).

5.7 Bonferroni correction and the inflation of alpha errors

Whether or not predictors are included in a linear or multilevel model is usually judged based on the size of the standardised regression coefficients and the corresponding p-value of the predictor. Usually, if $p < 0.05$, a predictor is assumed to contribute significantly to the model. However, handling a larger number of independent variables in this way might overfit the model as some predictors might enter the model only 'by chance'. To limit overfitting when using multiple significance tests, corrections such as the Bonferroni or Holm correction can be applied (Hox et al., 2017). The Bonferroni correction reduces the desirable p-values by dividing the targeted α-level by the number m of conducted significance tests: $\alpha_m = \alpha/m$. This method also translates to an altered meaning of asterisks when interpreting p-values, see Table 10, and can be applied for any occurrence of multiple comparisons.

Table 10. Symbols and meaning of p-values when applying Bonferroni correction.

Symbol	Usual meaning	Bonferroni correction to α-level
ns	$p > 0.05$	$p > 0.05/m$
*	$p \leq 0.05$	$p \leq 0.05/m$
**	$p \leq 0.01$	$p \leq 0.01/m$
***	$p \leq 0.001$	$p \leq 0.001/m$

5.8 Sampling and population

Although data samples are ideally analysed to draw conclusions with respect to a general population, this is often not possible in real-life settings because samples are rarely randomly selected to represent the whole population. Hence, the population of interest is sometimes best defined by the sample instead (Tabachnick & Fidell, 2013). Out-of-school learning opportunities often rely on the research sample provided by participants in their offers, who join because of their own motivation and not because of a random selection. Moreover, external constraints on group selection often interfere with a desired random selection. In this case, generalisations should be made carefully. S'Cool LAB is a special offer at CERN, only able to accommodate a marginal fraction of all students worldwide and a small fraction of the school groups coming to CERN for school trips. To inform the selection for S'Cool LAB programmes, the motivation of teachers and students is evaluated. In this way, particularly motivated groups are chosen because it is assumed that well-prepared students will benefit most from the activities. Finally, motivated groups are typically easier to handle for volunteering scientists and provide a more enjoyable experience for the volunteers.

Due to the external constraints, this procedure supports the **Matthew effect** (Merton, 1968), by offering even more activities to already very motivated students whose parents have the financial means to support a school trip to Switzerland. However, from a research perspective, the challenging sample provides a unique opportunity to test the effectiveness of an extreme example of an out-of-school learning lab. If S'Cool LAB were able to enhance the educational outcomes of motivated students even further, this would indicate a promising intervention also for other students.

5.9 Software and software parameters

The software package SPSS Statistics v25 was used for most of the statistical analysis. In particular, SPSS was used to screen and prepare the data (e.g., z-transformation), for principal axis factoring as well as for comparing group means through t-tests. Sequential linear regressions were performed through the option 'Analyze > Regression > Linear' while multi-level modelling used the option 'Analyze > Mixed Models > Linear' with the covariance type 'variance components'. Also, the restricted maximum likelihood (REML) estimation method was selected because it takes the number of estimated fixed effects into account (Harville, 1977).

However, SPSS does not support tetrachoric correlations. Here, STATA v13 was used to calculate tetrachoric correlation matrices with the command ‘tetrachoric [item_1]-[item_i], notable posdef; matrix C = r(corr)’ and to perform principal axis factoring with a predefined number of factors with the command ‘factormat C, n(…) factors(…)’, see also StataCorp (2019).

6 Pilot study – testing and developing questionnaires

From May to July 2016, a pilot study was conducted to test the effectiveness of the S'Cool LAB Day programme (see 4.4) and to examine the performance of different items and scales forming the pre-visit and post-visit questionnaires. Most of the items and scales had been developed and validated by other research groups in previous studies. However, they were adapted to the specific S'Cool LAB context, for example, by translating them into simple English. Here, section 6.1 describes the international student sample taking part in this pilot study. Section 6.2 provides an overview of the scales while section 6.3 summarises the item and instrument analysis. Furthermore, to assess students' conceptual understanding, several concept test items were tested, partially with free-response questions, to inform the development and selection of multiple-choice questions for the main study. This development is described in section 6.4. First results of the impact of the S'Cool LAB hands-on session are described in section 6.5 while section 6.6 summarises the learnings of the pilot study.

6.1 Study framework and sample

During the pilot study, students took part in the S'Cool LAB Day programme described in section 4.4, which included a guided tour and three hands-on workshops in S'Cool LAB, namely the electron tube, the cloud chamber, and X-ray workshop. Pre-visit questionnaires were administered in the morning of the S'Cool LAB Day right before the start of the guided tour. The post-visit questionnaires were administered through an online platform approximately 10 days after the S'Cool LAB Day programme (Figure 14). To connect the pre-visit questionnaires to the respective post-visit questionnaires, students generated an identification code based on the first letter of their first name, their mother's first name, their month of birth, and the number of older sisters and brothers.

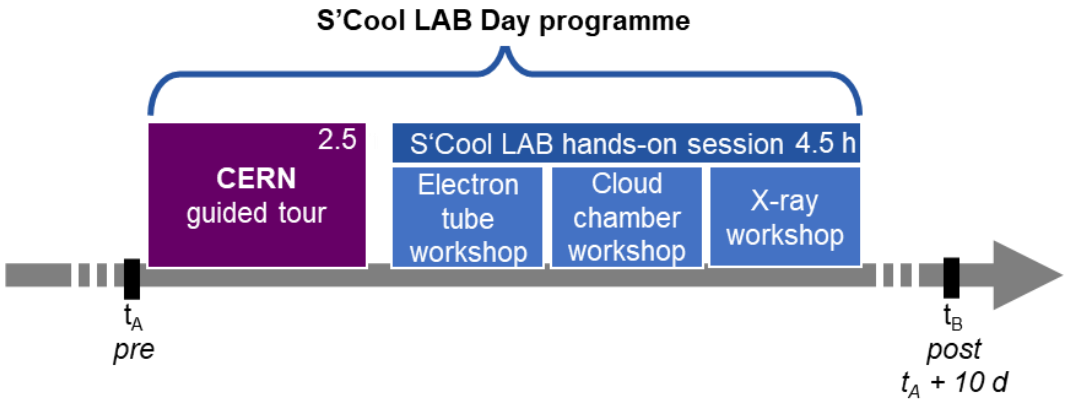


Figure 14. Timeline of the S'Cool LAB Day programme and questionnaires for the pilot study.

In total, four groups from three different countries (Germany, Turkey and Sweden) filled out the pre-visit questionnaire of the pilot study with a total of 106 students, 33% of whom were female. The average age of the students was $16.7\,y$ ($SD = 1.4\,y$). After taking part in the hands-on session in S'Cool LAB, 60 students filled out the post-visit questionnaire. Here, appendix 10.6 includes a more detailed overview of the sample of the pilot study. Due to technical constraints, one of the four student groups filled out the post-visit questionnaire directly at the end of the day at CERN. Therefore, certain scales of these questionnaires were not included in the analysis because some of the items are not sensitive at a direct evaluation after the S'Cool LAB hands-on session, for example, *"I spent time on physics I learned about in S'Cool LAB in my free time – even beyond homework."*.

6.2 Overview of items and scales forming the questionnaires

To measure the educational outcomes of the hands-on session in S'Cool LAB, as well as the novelty factors associated with students' S'Cool LAB experience, scales have been adapted from already existing instruments. These scales were chosen because previous studies confirmed favourable item and scale properties, such as internal consistency and validity. However, most of these scales had been used only in German in previous studies, and in slightly different contexts. Therefore, German items were translated into English by the author together with a bilingual colleague who assured the validity of the translations. In addition, the vocabulary of the items was simplified to reach a B1 level using the 'text inspector' developed by Bax (2016).

Most items included a statement such as *"I enjoy my physics classes"* and a six-level Likert-type rating scale coded from 1 to 6: completely disagree (1), disagree (2), disagree a little (3), agree a little (4), agree (5), completely agree (6). Furthermore, demographic information about the students and their relevant prior experiences were collected. In addition to students' educational outcomes, several items were included to evaluate additional aspects of the S'Cool LAB project including students' satisfaction with the workshops, with the room itself, and with their whole CERN experience. Although the latter items did not contribute directly to answering the research questions, they provided important feedback as part of the formative assessment of S'Cool LAB.

Table 11 shows an overview of the scales, including their origin, item number and internal consistency during the pilot study.

Table 11. Overview of items and scales used in the pilot study including their source and original language, if applicable. In addition, the item number and internal consistency α_C are listed as well as the number of items that were selected for the main study together with the internal consistency α_C for the reduced scale.

Construct	Source	original language	# items during the pilot study	Cronbach's α_C	# selected items for main study	Cronbach's α_C reduced scale
Demographic information & control variables (pre-visit questionnaire)						
Identification code	(Grube, Morgan, & Kearney, 1989)		1		1	
Age			1		1	
Gender			1		1	
Grade			1		1	
Native language			1		1	
Nationality			1		1	
Country of school			1		1	
Name of school			1		1	
Name of teacher			1		1	
Date of visit			1		1	
Favourite school subjects	(Zwick & Renn, 2000)	DE	4		4	
Novelty influence factors (pre-visit questionnaire)						
School grades						
Mathematics			1		1	
Physics			1		1	
English			1		1	
Prior experiences with …						
… physics classes			2		2	
… particle physics			3		3	
… experiments	(Börlin, 2012)	DE	4		4	
… out-of-school learning	(Cors, 2016)	DE	6		6	
Novelty experience factors (post-visit questionnaire)						
Perceived novelty	(Molz, 2016)	DE	14	0.63	14	0.63
Cognitive load						
Cogn. load experiments		DE	5	0.87	5	0.87
Cogn. load cloud chamber	(Molz, 2016)	DE	5	0.88	4	0.76[a]
Cogn. load electron tube		DE	5	0.91	4	0.84[a]
Cogn. load X-ray		DE	5	0.93	4	0.87[a]
Cognitive load NASA TLX	(Hart, 2006)	EN	5	0.73	5	0.73
Tutor support	(Pawek, 2009; Yunker, 2010)	DE	6	0.85	6	0.85

[a] A new item was added, two items were deleted. α_C is based on 3 of the original items.

Construct	Source	original language	# items during the pilot study	Cronbach's α_C	# selected items for main study	Cronbach's α_C reduced scale
Educational outcomes and their dispositional counterparts if applicable						
Interest physics	(Kuhn, 2010)					
Dispositional (pre)		DE	9	0.90	6	0.84
Situational (post)		DE	9	0.88	6	0.87
Self-beliefs physics	(Kuhn, 2010)					
Dispositional (pre)		DE	9	0.90	5	0.88
Situational (post)		DE	9	0.87	5	0.85
Self-beliefs experiments	(Maiseyenka, 2014)					
Dispositional (pre)		DE	6	0.89	4	0.86
Situational (post)		DE	6	0.88	4	0.83
Curiosity	(Litman & Spielberger, 2003; Naylor, 1981)					
Curiosity trait (pre)		EN	7	0.86	5	0.87
State, particle physics (pre)		EN	4	0.90	3	0.90
State, particle physics (post)		EN	10	0.97	6	0.96
Enjoyment (post)	(Rennie, 1994)	EN	5	0.81	5	0.81
Easiness (post)		EN	5	0.82	5	0.82
Helpfulness (post)		EN	5	0.83	5	0.83
Cognitive activation (post)	(Hänze & Berger, 2007)	EN	7	0.75	5	0.75
Other questions, mostly for the formative assessment of S'Cool LAB (post)						
Satisfaction with …						
… guided tour at CERN			3		3	
… workshops			3		3	
… the whole trip to CERN			4		4	
… the different experiments			3		3	
... the laboratory room			6		1	
Perception of time on task	(Yunker, 2010)	DE	4		0	
Details workshop team			3		3	
Important aspects of the trip	(Pawek, 2009)	DE	7		7	
Cognitive links			3		0	
Recommendation to peers			1		1	
Difficulties due to languages			3		3	
Feedback for questionnaire			2		2	
Conceptual understanding						
7 multi-tier items (pre)	Development see 6.4		0		15	
Confidence levels (pre)			0		7	
7 multi-tier items (post)			0		15	
Confidence levels (post)			0		7	
		SUM	**207**	SUM	**209**	

Indeed, the goal of the pilot study was to test all items and scales in the S'Cool LAB context, to reduce the number of items per scale to decrease the test load, and to select suitable item groups for the main study. Here, Table 11 already includes the results of the pilot study, namely the number of items per scale that were selected for the main study and the internal consistency of the reduced scales.

The questionnaires in their print version are included in appendices 10.1 and 10.2. Here, the design and layout of the questionnaires followed the practical recommendations by Porst (2008). For example, students received written instructions about the goal, conditions, and different item formats before filling out the questionnaire. Every item block was introduced by a short description to help participants transition between different parts of the questionnaire. On the last page, free space allowed participants to leave comments. To evaluate the duration of the questionnaires, a small group of native English-speaking students ($n = 17$) with an average age of 17.5 years noted down the time when starting the questionnaire and at the end of every page of the questionnaire. On average, students took 9 minutes to answer the 59 items of the pre-visit questionnaire and 18 minutes for the 127 items of the post-visit questionnaire.

6.3 Item and instrument analysis

The following section reports the properties of items and scales used during the pilot study and describes the elimination process that informed the selection of suitable item groups for the main study. The goal of the elimination analysis was to reduce all scales to approximately five items to reduce the test load on students. First, an item was eliminated if the scale's internal consistency measure Cronbach's α_C increased when omitting this item. Second, an item was considered for elimination if Cronbach's α_C did not decrease when eliminating this item. Here, the item text and its meaning with respect to the underlying theoretical construct were examined carefully before eliminating an item.

In addition, the item discrimination was evaluated through the corrected item-to-total correlation r_{tj}. Here, Ding and Beichner (2009) recommend a relatively low threshold of $r_{tj} > 0.2$. However, a specific item was also considered for removal if its corrected item-to-total correlation was significantly lower than that of other items in the same scale. Since the sample size of the pilot study was rather small, factor analysis was not performed. In the following, a short description of all items is given. The full version of all items is included in appendices 10.1 (pre-visit questionnaire) and 10.2. (post-visit questionnaire).

6.3.1 Dispositional and situational physics interest and self-beliefs

The basis of the interest and self-beliefs scales used in this study was the motivation test of the MARKUS Study (Helmke, 2002). These items were adapted and validated by Kuhn (2010) resulting in 26 six-level items separated in three motivation subscales: self-concept, interest and engagement, relationship to reality & authenticity. Molz (2016) and Hochberg (2016) further selected and adapted suitable items to assess students' interest and self-beliefs in the context of out-of-school learning labs before, during, and after taking part in an intervention. For the pilot study, these items were translated into English and adapted to the S'Cool LAB context.

Interest

The pre- and post-visit questionnaires of the pilot study included nine six-level items. Students rated their agreement with these nine statements from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). While the items in the pre-visit questionnaire focused on students' dispositional interest (e.g., *"I enjoy solving physics problems."*), the items in the post-visit questionnaire focused on students' situational interest related to the activities in S'Cool LAB (e.g., *"During my time in S'Cool LAB, I enjoyed solving physics problems."*).

All items displayed acceptable properties, and the scale showed a good internal consistency both in the pre and post-visit questionnaire, see Table 12.

However, to reduce the test load, item IE5 was discarded for the main study, because it was not possible to find suitable parallel items for the pre- and post-visit questionnaires. Furthermore, item IE2 was discarded because of its low corrected item-to-total correlation in the post-visit questionnaire $(r_{tj} = 0.28)$ and because Cronbach's α_C for the scale improved when omitting this item. Item IE6 was discarded because of its similarity to item IE1. The remaining items represent several components of interest, such as its epistemic component (IE1, IE9), its feeling-oriented component (IE3, IE8) and its social or value-oriented component (IE4), as well as the characteristic flow (IE7). The reduced scale shows good internal consistency with Cronbach's $\alpha_C = 0.84/0.87$ in the pre- and post-visit questionnaires.

Table 12. Item and scale properties of dispositional and situational interest rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval. An asterisk marks items that were discarded for the main study.

Physics interest		**Pre (n = 106)**				**Post (n = 42)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
IE1	I spend time on physics in free time	3.30	1.45	0.69	0.88	3.90	1.57	0.70	0.86
*IE2	I work harder in physics than other subjects	3.70	1.53	0.65	0.89	4.12	1.48	**0.28**	**0.90**
IE3	I enjoy solving physics problems	4.01	1.45	0.74	0.88	5.24	0.79	0.65	0.87
IE4	I discuss physics with friends or family	3.78	1.56	0.51	0.90	4.60	1.27	0.72	0.86
*IE5	Physics is one of my favourite subjects	3.88	1.80	0.83	0.87	4.19	1.64	0.59	0.87
*IE6	I learn more about physics at home	3.79	1.47	0.54	0.89	4.31	1.77	0.78	0.85
IE7	I do not notice time passing	3.51	1.55	0.61	0.89	4.88	1.23	0.70	0.86
IE8	I like physics classes/S'Cool LAB	4.25	1.41	0.54	0.89	5.33	0.85	0.61	0.87
IE9	I want to spend more time on physics/S'Cool LAB	3.82	1.62	0.79	0.87	5.10	1.14	0.85	0.85
	Cronbach's α_C (all items)	**0.90** (95% CI [0.86, 0.92])				**0.88** (95% CI [0.82, 0.93])			
	Cronbach's α_C (reduced item set, without *items)	**0.84** (95% CI [0.79, 0.89])				**0.87** (95% CI: [0.80, 0.92]			

Self-beliefs

The pre- and post-visit questionnaires included 15 six-level statements that had to be rated from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). While the items in the pre-visit questionnaire focused on students' dispositional self-beliefs with respect to their physics classes (e.g., *"I find it easy to solve tasks in physics classes."*), the items in the post-visit questionnaire focused on students' situational self-beliefs related to the activities in S'Cool LAB (e.g., *"I found it easy to solve tasks in S'Cool LAB."*). Five of the 15 items were added to measure students' self-beliefs related to doing physics experiments in more detail. Since S'Cool LAB welcomes students from many different countries with different physics curricula and education traditions, it was expected that students' prior experience in doing experiments would differ significantly with some students having only minimal hands-on experience. Here, Maiseyenka (2014) adapted items designed to measure students' physics-related self-concept from Brell (2008) to measure students' self-concept related to experiments (five items, Cronbach's $\alpha_C = 0.70$). These items were translated from German into English and adapted to the S'Cool LAB setting. Again, all items were used with a slightly different focus in the pre-visit questionnaire (e.g., *"Doing experiments in physics is easy for me."*) and the post-visit questionnaire (e.g., *"Doing experiments in S'Cool LAB was easy for me."*).

All items demonstrated acceptable properties, and the scale showed good internal consistency both in the pre and post-visit questionnaire, see Table 13 (physics self-beliefs) and Table 14 (experiment self-beliefs).

Table 13. Item and scale properties of dispositional and situational physics self-beliefs rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval. An asterisk marks items that were discarded for the main study.

Physics self-beliefs		Pre (n = 106)				Post (n = 60)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
*SC1	Physics classes/my time in S'Cool LAB is fun	4.16	1.44	0.61	0.88	5.40	0.83	0.62	0.85
SC2	I understand the learning material	4.36	1.29	0.69	0.88	5.22	0.99	0.67	0.85
SC3	I look forward to physics classes/ my time in S'Cool LAB	3.67	1.41	0.76	0.87	5.05	1.03	0.71	0.84
*SC4	I expect that my performance will be good in the future	4.12	1.57	0.76	0.87	4.68	1.32	**0.37**	**0.88**
SC5	My performance in physics classes/S'Cool LAB is good	3.98	1.53	0.69	0.88	5.00	0.96	0.81	0.83
*SC6	I actively take part in physics classes/S'Cool LAB	4.34	1.23	0.51	0.89	5.25	0.68	0.61	0.86
SC7	I find it easy to solve tasks	3.72	1.40	0.70	0.88	4.85	0.94	0.67	0.85
*SC8	I am focused during physics/SL	4.48	1.31	0.55	0.89	5.12	0.85	0.52	0.86
SC10	My classmates think I am good in physics	3.85	1.58	0.72	0.87	4.45	1.31	0.63	0.85
	Cronbach's α_C (all items)	**0.90** (95% CI [0.87, 0.93])				**0.87** (95% CI [0.81, 0.91])			
	Cronbach's α_C (reduced item set, without *items)	**0.88** (95% CI [0.83, 0.91])				**0.85** (95% CI [0.78, 0.90])			

However, to reduce the test load, several items were discarded for the main study. Item SC4 was omitted because it was not possible to find suitable parallel items for the pre- and post-visit questionnaire, which is also reflected in the comparatively low corrected item-to-total correlation of the post-visit item ($r_{tj} = 0.37$) and an improved consistency of the post-visit scale when removing this item.

Items SC6 and SC8 were discarded because they focused on aspects of cognitive activation, which is measured within a separate scale (see 6.3.6). Finally, item SC1 was discarded because it was closely related to the feeling-oriented component of interest instead of self-beliefs, which was also confirmed by the results of a factor analysis by Hochberg (2016). The remaining items represent several components of self-beliefs, such as students' perception of their performance in physics (SC2, SC5), comparison with other students (SC10) and their perceived competence feeling (SC7).

Table 14. Item and scale properties of dispositional and situational experiment self-beliefs rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval. An asterisk marks items that were discarded for the main study.

Experiment self-beliefs		Pre (n = 103)				Post (n = 59)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
SC9	Experiments help me to understand the learning material	4.88	1.15	0.59	0.89	5.08	0.95	0.70	0.86
*SCE1	Physics experiments are fun	4.97	1.12	0.60	0.88	5.31	0.77	0.59	0.88
SCE2	Doing experiments is easy for me	4.58	1.20	0.80	0.85	4.98	0.90	0.72	0.86
SCE3	I am good in doing experiments	4.49	1.22	0.86	0.84	5.02	0.90	0.70	0.86
SCE4	I find it easy to finish in time	4.36	1.26	0.69	0.87	5.12	0.93	0.68	0.86
*SCE5	I find it easy to solve tasks related to physics experiments	4.11	1.35	0.70	0.87	4.78	1.19	0.79	0.85
	Cronbach's α_C (all items)	**0.89** (95% CI [0.85, 0.92])				**0.88** (95% CI [0.83, 0.92])			
	Cronbach's α_C (reduced item set, without *items)	**0.86** (95% CI [0.81, 0.90])				**0.83** (95% CI [0.74, 0.89])			

Furthermore, Items SCE1 and SCE5 were discarded from the subscale self-beliefs related to physics experiments. Here, item SCE1 referred rather to the feeling-oriented aspect of interest than to self-beliefs, while item SCE5 was dropped because this item was very similar to item SCE2.

In summary, both reduced self-beliefs scales showed good internal consistency in the pre- and post-visit questionnaires with Cronbach's $\alpha_C > 0.83$.

6.3.2 Curiosity trait and curiosity state towards particle physics

Items based on Litman and Spielberger (2003) and Naylor (1981) were adapted to and tested in out-of-school learning labs by previous studies in Germany (Hirth, 2019; Hochberg, 2016; Molz, 2016). For this study, items were used in their original version in English or translated from German into English and adapted to the S'Cool LAB context.

The pre-visit questionnaires included seven six-level items to measure students' dispositional curiosity (curiosity trait, e.g., *"I find it fascinating to learn new things."*) as well as four six-level items to measure their curiosity towards particle physics (curiosity state, e.g., *"I find it fascinating to spend time on particle physics."*). Moreover, the post-visit questionnaire included nine six-level items to measure students' curiosity towards particle physics (curiosity state, e.g., *"The time in S'Cool LAB made me curious to learn more about particle physics."*) after taking part in the S'Cool LAB hands-on session. Students rated their agreement with these statements from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6).

Curiosity trait

Although all seven items showed acceptable properties and the scale showed good internal consistency, the items displayed a tendency towards a ceiling effect with mean values well above 5 ($\cong 80\%$ of the maximum), see Table 15. Nevertheless, suitable items were selected for the main study. To reduce the test load, items CT3 and CT7 were discarded for the main study because of their comparatively low corrected item-to-total correlation ($r_{tj} = 0.47/0.56$) and because the internal consistency of the scale did not decrease when omitting these items. The reduced scale showed good internal consistency with $\alpha_C = 0.87$.

Table 15. Item and scale properties of curiosity trait rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval. An asterisk marks items that were discarded for the main study.

Curiosity trait		**Pre (n = 105)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CT1	I find it fascinating to learn new things	5.32	0.78	0.66	0.84
CT2	I am curious about new things	5.28	0.83	0.76	0.82
*CT3	Critical thinking about previous ideas is fun	4.81	0.91	**0.47**	**0.86**
CT4	I like to learn things that are new to me	5.28	0.84	0.69	0.83
CT5	I find new things inspiring	5.12	0.90	0.62	0.84
CT6	When I hear something new, I want to know more	5.07	0.92	0.74	0.82
*CT7	I enjoy studying things that I don't understand	4.52	1.23	**0.56**	**0.86**
	Cronbach's α_C (all items)	**0.86** (95% CI [0.82, 0.90])			
	Cronbach's α_C (reduced item set, without *items)	**0.87** (95% CI [0.83, 0.91])			

Curiosity state towards particle physics

All items demonstrated acceptable properties, and the scale showed good internal consistency, see Table 16. However, items CS3, CS4 and CS6 were discarded for the main study, because the internal consistency of the scale did not decrease considerably when omitting these items. Consequently, three items remained in the pre-visit questionnaire with $\alpha_C = 0.90$ and six items remained in the post-visit questionnaire with $\alpha_C = 0.96$.

Both the number and the focus of some of the items differed between the pre- and post-visit questionnaire because some items referred to the current state of curiosity, such as CS9 *"I find it fascinating to spend time on particle physics."* while other items focused on the new curiosity that was aroused by activities in S'Cool LAB, such as CS1 *"My time in S'Cool LAB drew my attention drawn to the topic of particle physics"*. Therefore, pre-post comparisons should be interpreted cautiously.

Table 16. Item and scale properties of curiosity state rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval. An asterisk marks items that were discarded for the main study.

Curiosity state		Pre (n = 103)				Post (n = 60)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
CS1	My time in S'Cool LAB drew my attention drawn to the topic					4.82	1.32	0.85	0.96
CS2	I want to learn more about particle physics	4.86	1.07	0.84	0.86	4.60	1.36	0.90	0.96
*CS3	I feel/felt a personal connection	3.96	1.20	0.72	**0.90**	4.45	1.40	0.84	0.96
*CS4	S'Cool LAB allowed me to study phenomena I didn't know before					4.70	1.18	0.65	**0.97**
CS5	S'Cool LAB made me curious to learn more about particle physics					4.72	1.34	0.89	0.96
*CS6	The experiments were inspiring					4.72	1.29	0.91	0.96
CS7	I am more curious now					4.47	1.43	0.88	0.96
CS8	I want to understand the topic better	4.90	1.04	0.77	0.88	4.68	1.40	0.88	0.96
CS9	I find it fascinating to spend time on particle physics	4.40	1.26	0.82	0.86	4.50	1.44	0.92	0.96
	Cronbach's α_C (all items)	**0.90** (95% CI [0.87, 0.93])				**0.97** (95% CI [0.95, 0.98])			
	Cronbach's α_C (reduced item set, without *items)	**0.90** (95% CI [0.86, 0.93])				**0.96** (95% CI: [0.95, 0.98])			

6.3.3 Enjoyment, helpfulness, and easiness

To assess students' perceived enjoyment, helpfulness and easiness related to the hands-on session in S'Cool LAB, items by Rennie (1994) were adapted to the S'Cool LAB context. Instead of referring to 'working in groups' the items texts were adjusted to 'doing the experiments'. Furthermore, all items were adapted to match the response options of the other items in the questionnaire to make it easier and faster for students to reply. In particular, the post-visit questionnaires included five six-level items for each of the three subscales and students rated their agreement with five statements each from 'disagree completely' (coded as 1) to 'agree completely' (coded as 6).

Enjoyment

All five items showed acceptable properties, and the scale displayed good internal consistency with $\alpha_C = 0.81$, see Table 17. Only item EN4 showed a comparatively low corrected item-to-total correlation ($r_{tj} = 0.43$) and omitting this item would have improved the scale. However, due to the limited sample size of the pilot study and the limited number of items in this scale, all items were kept for the main study to observe their behaviour in a larger sample.

Table 17. Item and scale properties of perceived enjoyment rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

Enjoyment		Post (n = 60)			
Label	Item text	M	SD	r_{tj}	α_{tj}
EN1	I enjoyed using the equipment.	5.38	0.64	0.65	0.76
EN2	I enjoyed preparing the experiments.	5.18	0.81	0.69	0.74
EN3	I enjoyed doing the experiments.	5.45	0.77	0.65	0.75
EN4	I enjoyed discussing the phenomena.	5.00	1.06	**0.43**	**0.84**
EN5	I enjoyed the time in S'Cool LAB.	5.43	0.74	0.65	0.75
	Cronbach's α_C	**0.81** (95% CI [0.72, 0.87])			

Helpfulness

All five items displayed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.83$, see Table 18. Thus, all items were kept for the main study.

Table 18. Item and scale properties of perceived helpfulness rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

Helpfulness		Post (n = 60)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}
HE1	S'Cool LAB was helpful for my physics classes	4.25	1.34	0.62	0.80
HE2	It helped me to get better at doing experiments	4.82	1.03	0.65	0.79
HE3	It helped me to understand physics concepts	4.97	0.90	0.71	0.78
HE4	It helped me get an idea how scientists work	4.97	1.06	0.62	0.79
HE5	It helped me get an idea about what CERN does	4.87	1.27	0.57	0.81
	Cronbach's α_C	**0.83** (95% CI [0.75, 0.89])			

Easiness

All five items demonstrated acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.82$, see Table 19. Only item EA5 showed a comparatively low corrected item-to-total correlation ($r_{tj} = 0.49$) and omitting this item would have improved the scale. However, all items were kept for the main study to observe their behaviour in a larger sample size.

Table 19. Item and scale properties of perceived easiness rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

Easiness		Post (n = 60)			
Label	Item text	M	SD	r_{tj}	α_{tj}
EA1	It was easy to understand the instructions.	4.80	0.90	0.50	0.82
EA2	It was easy to prepare the experiments.	4.87	0.85	0.74	0.75
EA3	It was easy to use the equipment.	5.07	0.80	0.75	0.75
EA4	It was easy to get a result.	4.95	0.93	0.63	0.78
EA5	It was easy to understand what it was all about.	4.73	0.99	**0.49**	**0.83**
	Cronbach's α_C	**0.82** (95% CI [0.74, 0.88])			

6.3.4 Novelty experience

Molz (2016) developed an instrument to assess students' perceived novelty in the context of an out-of-school hands-on learning lab based on existing items from the SOLEI (Science Outdoor Learning Environment Inventory) instrument (Orion, Hofstein, Tamir, & Giddings, 1997). In particular, he adapted items to capture the effects of preparation aiming at reducing students' perceived novelty onsite. Factor analysis confirmed four factors, which Molz (2016) interpreted as

- 'setting orientation' (e.g., *"Thanks to the preparation, the visit to the student lab was well organised."*)
- 'capabilities' (e.g., *"Thanks to the preparation, we received enough information to conduct the experiments."*)
- 'fun' (e.g., *"Thanks to the preparation, it was fun to try out the experiments."*)
- 'lack of fear' (e.g., *"Thanks to the preparation, I was not afraid of making mistakes during the experiments."*)

For this study, suitable items were translated from German into English unless the original items were available in English. Furthermore, the scales were extended using additional items from the SOLEI instrument (Orion et al., 1997; Yunker, 2010), such as *"The materials were easy to find."*. All item texts were simplified to a B1 English level. In addition, reverse wording was resolved to avoid confusion and to make the items easier to understand as suggested, for example, by Van Sonderen, Sanderman, and Coyne (2013). For example, "*I was not afraid of …"* was transformed into *"I was afraid of …"*. Consequently, negatively worded items were coded inversely. Thus, the post-visit questionnaires included 14 six-level items and students rated their agreement with these statements from 'disagree completely' to 'agree completely'.

Five of the items showed problematic properties with respect to a common novelty scale, which is reflected in a relatively low internal consistency with $\alpha_C = 0.63$, see Table 20.

The original complete novelty scale by Molz (2016) showed a much higher internal consistency with $\alpha_C = 0.91$, despite its underlying factor structure. However, the S'Cool LAB setting differed significantly from the setting in his studies. In particular, students coming to S'Cool LAB did not receive specific preparation aimed at reducing setting novelty. Therefore, items were expected to behave differently and to form different factors. Nonetheless, since the sample size of the pilot study was rather small, all items were kept for the main study and factor analysis was postponed to the main study.

Table 20. Item and scale properties of perceived novelty rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

Novelty		Post (n = 58)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}
TOP1	I was familiar with the schedule in S'Cool LAB.	4.60	1.31	0.32	0.60
TOP2	The time in S'Cool LAB was well organized.	5.47	0.60	0.45	0.60
TOP3	I was well prepared for my time in S'Cool LAB.	4.74	1.22	0.39	0.59
TOP4	I knew what to expect in S'Cool LAB.	4.02	1.43	0.31	0.60
TOP5	I was easily getting familiar with the work places	4.95	1.02	0.56	0.57
TOP6	Before arriving, I knew how I was going to work	3.62	1.78	**0.15**	**0.64**
TOP7	The materials needed were easy to find.	5.14	1.15	0.37	0.59
TIN1-	It was hard for me to do the experiments. (inverse)	4.72	1.36	**0.14**	**0.63**
TIN2-	I was afraid to make mistakes (inverse)	3.86	1.59	**0.09**	**0.65**
TIN3-	It took a long time until I was familiar with the equipment & able to start the experiments (inverse)	4.28	1.42	**0.05**	**0.65**
TIN4	It was fun to try out the experiments.	5.41	0.75	0.37	0.60
TIN5	I received enough information to do experiments	5.21	0.74	0.46	0.59
TIN6	I was able to concentrate on the experiments, without "fighting" with the equipment.	5.19	0.80	0.37	0.60
TIN7	The instructions were helpful for doing experiments	5.33	0.89	**0.17**	0.62
	Cronbach's α_C (all items)	**0.63** (95% CI [0.47, 0.76])			

6.3.5 Cognitive load

The workshops in S'Cool LAB offer different levels of complexity related to the equipment used. Therefore, the cognitive load scale by Hirth (2019) was adapted for the S'Cool LAB context by focusing on the use of the equipment in the different workshops and by inverting some items to ensure that all items of the scale measured in 'the same direction'. The last step was necessary to simplify the questionnaire for non-native speaking users. In total, 25 items distributed over five subscales were included in the post-visit questionnaire to assess students' perceived cognitive load.

Cognitive load of the experiments in general

Five items referred to the cognitive load of the experiments in general, see Table 21, with items focusing, for example, on the perceived difficulty *("The experiments were difficult."*) but also to the germane cognitive load of the underlying physics principles (*"I had problems understanding the physics principles of the experiments."*) and the extraneous cognitive load offered by the worksheets *("It was hard for me to understand the instructions."*). Students rated their agreement with these statements from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6).

Table 21. Item and scale properties of perceived cognitive load in general rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

Cognitive load experiments in general		Post (n = 60)			
Label	**Item text**	M	SD	r_{tj}	α_{tj}
CLE1	I had problems interpreting my observations correctly.	2.83	1.36	0.73	0.83
CLE2	The experiments were difficult.	2.33	1.28	**0.51**	**0.88**
CLE3	I had problems understanding the physics principles of the experiments.	2.87	1.46	0.71	0.83
CLE4	It was hard for me to understand the instructions.	2.33	1.26	0.70	0.84
CLE5	I would have needed more assistance when working with the experiments.	2.38	1.32	0.81	0.81
	Cronbach's α_C (all items)	**0.87** (95% CI [0.81, 0.91])			

All five items displayed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.87$, see Table 21. Only item CLE2 showed a comparatively low corrected item-to-total correlation $(r_{tj} = 0.51)$ and omitting this item would have improved the scale slightly. Nevertheless, all items were kept for the main study because the number of items forming this scale was already relatively low.

Cognitive load of the different workshops

Five items each referred to the cognitive load of the different workshops (i.e., the cloud chamber workshop, the electron tube workshop and the X-ray workshop), see Table 22. These items were analysed mainly as part of the formative assessment of the S'Cool LAB activities.

For example, items referred to the perceived difficulty of using the equipment *("It was difficult to use the equipment when doing the experiment."*), the feeling of 'cognitive overload' *("I was overwhelmed by the features provided by the equipment."*) or to barriers towards the experimentation *("I had problems to start working with the equipment."*). All 15 items demonstrated acceptable properties, and the three subscales each showed good internal consistency with $\alpha_C = 0.88 - 0.93$, see Table 22.

However, to reduce the test load, the second item of every subscale (e.g., CLC2) was discarded for the main study, because it was very similar to the first item of every subscale (e.g., CLC1). Furthermore, the third item of every subscale (e.g., CLC3) was discarded due to its similarity with the fifths item (e.g., CLC5). The reduced scales still showed good internal consistency with a Cronbach's $\alpha_C = 0.76 - 0.87$. However, the remaining items focused mainly on the extraneous cognitive load provided by the equipment. Hence, one item was added to every subscale to evaluate the cognitive load required for understanding

the underlying physics concepts of every workshop (e.g., CLC6 "*I had problems understanding the physics principles of this experiment"*). This item was already part of the subscale 'cognitive load of the experiments in general' (CLE3) and demonstrated suitable properties.

Table 22. Item and scale properties of perceived cognitive load of the different workshops on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval. An asterisk marks items that were discarded for the main study.

Cognitive load of the cloud chamber workshop		**Post (n = 60)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CLC1	It was difficult to use the equipment.	2.08	1.09	0.69	0.86
*CLC2	It was difficult to use equipment when doing experiments	2.02	1.07	0.78	0.85
*CLC3	I had problems to start working with the equipment.	1.98	1.07	0.78	0.84
CLC4	I was overwhelmed by features of the equipment	2.07	1.18	0.72	0.86
CLC5	I had problems with the equipment making observations	2.15	1.19	0.65	0.88
	Cronbach's α_C (all items)	**0.88** (95% CI [0.83, 0.92])			
	Cronbach's α_C (reduced item set, without *items)	**0.76** (95% CI [0.63, 0.85])			

Cognitive load of the electron tube workshop		**Post (n = 42)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CLEl1	It was difficult to use the equipment.	2.31	1.22	0.80	0.89
*CLEl2	It was difficult to use equipment when doing experiments	2.43	1.35	0.78	0.89
*CLEl3	I had problems to start working with the equipment.	2.38	1.34	0.78	0.89
CLEl4	I was overwhelmed by features of the equipment	2.38	1.23	0.72	0.90
CLEl5	I had problems with the equipment making observations	2.24	1.25	0.80	0.89
	Cronbach's α_C (all items)	**0.91** (95% CI [0.86, 0.95])			
	Cronbach's α_C (reduced item set, without *items)	**0.84** (95% CI [0.73, 0.91])			

Cognitive load of the X-ray workshop		**Post (n = 42)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CLX1	It was difficult to use the equipment.	2.05	1.08	0.83	0.92
*CLX2	It was difficult to use equipment when doing experiments	2.17	1.27	0.81	0.92
*CLX3	I had problems to start working with the equipment.	2.43	1.43	0.86	0.91
CLX4	I was overwhelmed by features of the equipment	2.40	1.42	0.81	0.92
CLX5	I had problems with the equipment making observations	2.17	1.31	0.81	0.92
	Cronbach's α_C (all items)	**0.93** (95% CI [0.89, 0.96])			
	Cronbach's α_C (reduced item set, without *items)	**0.87** (95% CI [0.78, 0.93])			

NASA Task Load Index

Five items were adapted from the NASA Task Load Index (NASA-TLX). Molz (2016) and Cors (2016) first adjusted this scale to fit the context of out-of-school learning labs. Adapted to the S'Cool LAB context, items referred to different dimensions of task load, for example, mental demand (*"How high was the level of mental stress you experienced …"*), effort (*"How hard did you have to work to perform the experiments?"*), frustration (*"How frustrated*

were you …?"), and temporal demand (*"How do you rate the duration of the different experiments?"*).

Students answered these five questions on a seven-level semantic differential that included descriptions only at the endpoints. The lower endpoint was coded as 1, the neutral point as 4, and the higher endpoint was coded as 7. All five items showed acceptable properties, and considering its multi-dimensionality, the scales displayed a surprisingly good internal consistency with $\alpha_C = 0.73$, see Table 23. However, the corrected item-to-total correlations were relatively low for three of the items. Nevertheless, the scale was kept for the main study to evaluate its properties within a larger sample.

Table 23. Item and scale properties of perceived task load on a seven-level semantic differential: Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

NASA TLX		Post (n = 60)			
Label	**Item text**	M	SD	r_{tj}	α_{tj}
CL1	How high was the level of mental stress you experienced in S'Cool LAB? (very low (1) - very high (7))	3.67	1.65	0.70	0.59
CL2	How hard did you have to work to perform the experiments? (very little(1) - very much (7))	3.93	1.60	**0.46**	0.70
CL3	How frustrated were you during your time in S'Cool LAB? (very little (1) - very much (7))	2.38	1.42	**0.40**	0.72
CL4	How do you rate the duration of the different experiments? (very short (1) - very long (7))	3.90	1.08	**0.42**	0.71
CL5	How do you rate the number of different experiments you were working with? (not enough (1) - too many (7))	3.58	1.42	0.51	0.68
	Cronbach's α_C (all items)	**0.73** (95% CI [0.61, 0.83])			

6.3.6 Cognitive activation

Five six-level items based on Hänze and Berger (2007) were adapted to the S'Cool LAB context to measure students' perceived cognitive activation in the post-visit questionnaire. The scale included two items about students' cognitive engagement (e.g., *"I was focused when doing the experiments in S'Cool LAB."*). The other items were derived from effective deep-level learning strategies, such as critical review (*"I took a critical look at new ideas or new material."*), structuring information *("I tried to distinguish between important and unimportant things..."*) and elaboration (*"I tried to connect what I was learning with things I already knew.").*

In addition, two items were added to examine the cognitive activation presumed to be caused by predict-observe explain tasks (*"I thought about what will happen during the experiment before starting.",* and *"I was able to use my physics knowledge to interpret my*

observations during the experiments."). Students rated their agreement with all seven statements from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6).

All items demonstrated acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.79$, see Table 24. However, to reduce the test load, the two newly added items CA6 and CA7 were discarded for the main study, as they were included mainly to inform the further development of the activities in S'Cool LAB. Also, item CA6 showed a comparatively low corrected item-to-total correlation and Cronbach's α_C did not decrease when omitting this item. The reduced scale still showed good internal consistency with Cronbach's $\alpha_C = 0.75$.

Table 24. Item and scale properties of perceived cognitive activation rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval. An asterisk marks items that have been discarded for the main study.

		Post (n = 59)			
Label	**Item text**	M	SD	r_{tj}	α_{tj}
CA1	I was focused when doing the experiments in S'Cool LAB.	5.10	0.80	0.51	0.77
CA2	I was actively involved in doing the experiments in S'Cool LAB.	5.36	0.76	0.52	0.77
CA3	I took a critical look at new ideas or material.	4.39	1.22	0.52	0.77
CA4	I tried to distinguish between important and unimportant things when doing the experiments.	4.59	0.98	0.55	0.76
CA5	I tried to connect what I was learning with things I already knew.	4.97	0.98	0.58	0.76
*CA6	I thought about what will happen during the experiment before starting.	4.64	1.12	**0.43**	**0.79**
*CA7	I was able to use my physics knowledge to interpret my observations during the experiments.	4.92	1.24	0.60	0.75
	Cronbach's α_C (all items)	**0.79** (95% CI [0.70, 0.86])			
	Cronbach's α_C (reduced item set, without *items)	**0.75** (95% CI [0.63, 0.84])			

6.3.7 Tutor support

Six six-level items based on Pawek (2009) and the SOLEI instrument (Orion et al., 1997; Yunker, 2010) were adapted to the S'Cool LAB context to measure students' perceived support by tutors and their perception of the learning atmosphere. The scale included, for example, items about the approachability of the tutors *("I had the chance to ask the tutors in S'Cool LAB questions."*), their fascination about physics *("I had the feeling that the tutors are fascinated by physics."*), and the support students received from tutors *("The tutors helped me with problems while doing the experiments."*). Students rated their agreement with all six statements from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6).

Although all six items demonstrated acceptable properties and the scale displayed good internal consistency with $\alpha_C = 0.85$, the items showed a strong tendency towards a ceiling effect with mean values well above 5 ($\cong 80\%$ of the maximum), see Table 25. However, due to the small sample size, all items were kept the same for the main study to observe their behaviour in a larger sample.

Table 25. Item and scale properties of perceived tutor support rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

Tutor support		**Post (n = 59)**			
Label	**Item text**	M	SD	r_{tj}	α_{tj}
TS1	I had the chance to ask the tutors in S'Cool LAB questions.	5.53	0.68	0.66	0.82
TS2	I had the feeling that the tutors are fascinated by physics.	5.63	0.64	0.58	0.83
TS3	The discussions with the tutors were important to me.	5.22	0.89	0.70	0.81
TS4	I liked the working atmosphere in S'Cool LAB.	5.56	0.68	0.63	0.82
TS5	The tutors responded to my questions and comments while doing experiments.	5.61	0.53	0.68	0.82
TS6	The tutors helped me with problems while doing the experiments.	5.69	0.50	0.62	0.83
	Cronbach's α_C (all items)	**0.85** (95% CI [0.78, 0.90])			

6.3.8 Additional items

In addition to the scales described above, several single items were included in the questionnaires, for example, to assess students' prior experiences with physics, hands-on experiments or out-of-school learning. These items were kept for the main study and are not reported here. The full item texts of all items in the pilot study are included in appendices 10.1 and 10.2.

Of immediate relevance are students' answers to the post-visit questions *"What do you think about the length of this questionnaire? (*very short(1) – very long(7))*"* and *"How difficult was it for you to understand the questions in this questionnaire? (very difficult(1) – very easy(7))".* The average value for students' perception of the length of the questionnaire was $M = 5.82$ $(SD = 1.07)$ on a seven-level scale, which corresponds to 80% of the scale maximum. Similarly, the average value for students' perception of the difficulty of the questionnaire (or rather 'easiness' due to the polarity of the scale) was $M = 5.82$ $(SD = 1.40)$, also corresponding to 80% of the scale maximum. In summary, although students perceived the questions in the questionnaire as easy to understand, they perceived the questionnaire as long.

6.4 Development of concept test items

One of the goals of the hands-on workshops in S'Cool LAB is to provide cognitive conflict through surprising observations if students hold common misconceptions. Consequently, a small number of concept-test items was adapted from existing test instruments or developed based on previously documented students' conceptions (see section 2.4). These items were included to assess students' conceptual learning during the hands-on session in S'Cool LAB in addition to other affective and cognitive outcomes. Here, validated multiple-choice items about students' conception of radiation, which were suitable to measure learning outcomes of the X-ray workshop, were adapted directly from previous studies. However, multiple-choice items covering selected learning goals of the electron tube workshop and the cloud chamber workshop were not available. Hence, free-response questions were used first with a small sample of students to develop suitable distractors based on the students' answers. Then, the most promising items were selected and integrated into the pre- and post-visit questionnaires of the main study, as described below.

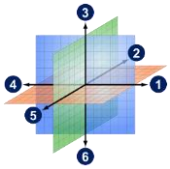
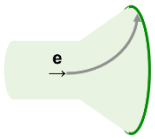
6.4.1 Electron tube workshop: the Lorentz force

Understanding the deflection of electrically charged particles in magnetic fields is a key concept of the electron tube workshop in S'Cool LAB. Furthermore, previous research suggests that students have difficulties with this topic because they confuse effects of electric and magnetic fields, e.g., by thinking of 'charged poles' (Maloney et al., 2001), see subsection 2.4.1. Concept inventories such as the 'Magnetism Conceptual Survey' (MCS) (Li & Singh, 2017) or the 'Conceptual Survey of Electricity and Magnetism' (CSEM) (Maloney et al., 2001) are available to test university-level students' understanding of magnetism. However, only a few of these items are accessible to high-school students.

Table 26 lists five multi-tier items that were adapted from existing test instruments and were tested with a small group of 58 German high-school students in July 2016 directly before and immediately after the S'Cool LAB hands-on session. Contrary to the existing test instruments, these items included several tiers to measure students' reasoning in more detail. Indeed, a study by Tamir (1989) "*confirmed the usefulness of justifications as a diagnostic tool and offered recommendations regarding future uses of justifications, including their contribution to the construction of two-tier items*." Although previous research offers several promising distractors for students' reasoning, as a first step, free-response questions were used. In this way, students could elaborate on their reasoning and their answers were used to derive suitable distractors for the main study. In four of the

five items (items PR, PS, PD, ED), students were first asked whether bar magnets exert a force on an electrically charged particle in a certain situation at tier-1 (answer options: yes or no). At tier-2, students were then asked about the direction of this force (multiple-choice, only relevant if "yes" was chosen at tier-1), while at tier-3, students were asked to explain their reasoning (free response). In contrast, item ET only asked for the direction of the magnetic field at tier-1 (multiple choice) and an explanation at tier-2 (free response).

Table 26. Item development for the electron tube workshop: the Lorentz force. The table lists the item labels, item descriptions including pictures as well as the item sources.

Item label, picture & source	Item text
PR Based on item 13 of MCS (Li & Singh, 2017)	A proton is moving to the right through the uniform magnetic field between two bar magnets.
PS Based on item 21 CSEM & item 5 MCS (Li & Singh, 2017; Maloney et al., 2001)	A proton is at rest in the uniform magnetic field between two bar magnets.
PD Based on (Scaife & Heckler, 2011)	A proton is moving downwards through the uniform magnetic field between two bar magnets.
ED Based on (Scaife & Heckler, 2011)	An electron is moving with constant speed downwards through the uniform magnetic field between two bar magnets.
PR, PS, PD, ED, ET 	**Questions for each of the four items above:** Tier-1: Does a force act on the proton/electron due to the bar magnets? (YES/NO) Tier-2: If so, in which direction? (multiple choice: 1-6, other) Tier-3: Please explain, why you think so! (free response)
ET  Based on item 22 of CSEM and item 23 of BEMA (Ding et al., 2006; Maloney et al., 2001)	An electron is moving horizontally towards a screen. The electron moves along the path that is shown in the picture because of a magnetic force caused by a magnetic field. Tier-1: In what direction does that magnetic field point? (multiple choice: 1-6, other) Tier-2: Please explain, why you think so! (free response)

After reading all answers of the free-response questions, students' answers were assigned to one of the following inductively formed reasoning categories:

- **C-RHR**: Correct explanation by correctly applying the right-hand-rule (RHR)
- **C180-RHR**: 180° shifted deflection angle, but otherwise correct application of RHR
- **W-RHR**: Wrong explanation by incorrectly applying the RHR
- **W-force||field**: Wrong explanation, 'force is parallel to magnetic field' conception
- **W-move=force**: Wrong explanation, 'movement of the particle indicates the direction of force' conception
- **W-charged**: Wrong explanation, 'charged poles' conception
- **W-balance**: Wrong explanation, 'forces balance out' conception
- **Other**: other explanations

Table 27 summarizes the results of the categorisation in the pre- as well as the post-test. In total, 154 written explanations distributed over four different items were analysed from the pre-test; 77 were analysed from the post-test.

Table 27. Results of the analysis of the free-response questions of items PR, PS, PD, ED, ET before and after participating in the S'Cool LAB hands-on session including the number of coded statements per item and explanation category, the total number (n) and percentage (%) of statements per category.

	Items pre-test							Items post-test						
Categories	**PR**	**PS**	**PD**	**ED**	**ET**	**n**	**%**	**PR**	**PS**	**PD**	**ED**	**ET**	**n**	**%**
C-RHR	13	20	12	8	5	58	**38%**	17	13	10	11	7	58	**75%**
C180-RHR	0	0	6	7	3	16	**10%**	0	0	4	3	1	8	**10%**
W-RHR	10	1	2	3	1	17	**11%**	1	0	0	1	4	6	**8%**
W-force||field	4	4	0	0	3	11	**7%**	0	0	0	0	0	0	**0%**
W-move=force	5	0	4	4	1	14	**9%**	0	0	0	0	0	0	**0%**
W-charged	9	7	5	7	0	28	**18%**	0	1	0	0	1	2	**3%**
W-balance	0	5	0	0	0	5	**3%**	0	3	0	0	0	3	**4%**
Other	2	2	1	0	0	5	**3%**	0	0	0	0	0	0	**0%**
#statements	**43**	**39**	**30**	**29**	**13**	**154**		**18**	**17**	**14**	**15**	**13**	**77**	

In summary, the following categories suggest promising distractors: 'the force is parallel to the magnetic field', 'the movement of the particle indicates the direction of the force' and the 'charged poles' conception, which is consistent with previous findings (Scaife & Heckler, 2011). Hence, the following answer options were included in the main study at tier-3 for the items PD, PR and PS to represent typical reasoning patterns:

- The south pole S of the magnets attracts the proton.
- The north pole N of the magnets attracts the proton.
- The magnetic force acts in the direction of the magnetic field.
- The magnetic force acts in the direction of the movement of the proton.
- The magnetic force acts on the proton according to the right-hand-rule.
- None of the explanations above apply.

6.4.2 Cloud chamber workshop: cosmic particles

During cloud chamber workshops in S'Cool LAB, students observe different types of particle tracks from natural radiation. Most of the tracks are caused by cosmic particles. In particular, secondary (anti)muons are produced in air showers when primary cosmic particles (mostly protons) interact with the atmosphere. Also, 'tertiary' electrons and positrons are produced when (anti-) muons transform at the end of their short lifetime.

Furthermore, terrestric radiation can be observed. For example, radioactive alpha-decays of Radon isotopes in the air cause the emission of alphas, which create easily visible tracks in the cloud chamber. In addition, due to the decay chain of these isotopes, electrons are emitted and their tracks can also be observed.

Previous research results suggest that many students are unaware of cosmic particles (see subsection 2.4.2). Consequently, many students may hold the misconception that the Earth, due to its radioactive rocks, is the main source of ionising particles that can be detected by particle detectors. Indeed, this conception was prevalent among physicists in the early days of particle physics before the famous balloon flights with a particle detector by Victor Hess (Physics Nobel Prize, 1936) and hiking expeditions, for example, by Anderson and Neddermeyer (1936). However, this conception had not been identified among students before nor do standardised test instruments exist for this topic.

Therefore, four multi-tier items were developed based on these historical misconceptions and historic experiments, see Table 28. Two of the items asked students to compare the detection rates of two particle detectors placed on Earth at $0\,m$ altitude and at an altitute of $5000\,m$ (item CB: detector on a balloon flight and item CM: detector on a mountain). Item CO asked students to compare the detection rates of two particle detectors placed either on the Moon's surface or on the Earth's surface. Finally, item CS asked students to compare detectors placed on the Earth's surface and on the International Space Station. All four items included two tiers: Tier-1 involved a yes-no question about the comparison of the detection count rate, while tier-2 involved a free-response question to explain students' reasoning.

Again, 58 German high-school students filled out these items before and after the S'Cool LAB hands-on session. After reading all free-text explanations of these four items, the answers of 58 students were assigned to one of inductively formed categories:

- **Shielding**: The Earth's atmosphere shields us from cosmic radiation.
- **Same atmosphere**: If both detectors are placed in the same atmosphere of the Earth, there will be the same number of (air) particles because there is a constant number of particles everywhere.
- **Space = source**: The source of the detected particles is space. The closer the detector is to the source, the more particles it will detect.
- **Cosmic particles**: Cosmic particles are responsible for the different detection rate.
- **Fewer air particles**: If there are fewer air molecules on the mountain or in space, the detector will measure fewer particles.
- **Earth = source**: The Earth is the only source of the detected particles. The further away the detector is from the source, the fewer particles will be detected.
- **Magnetic shielding**: The Earth's magnetic field has a shielding effect.
- **Other**: other explanations

Table 28. Item development for cloud chamber workshop: cosmic particles. The table lists the item labels, item descriptions and pictures.

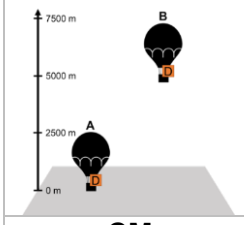
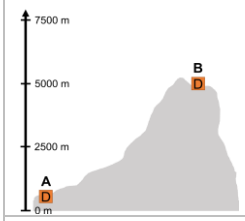
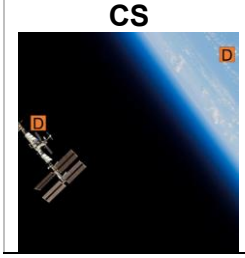
Label & picture	Item text
CB 	**Balloon flights with a particle detector** Imagine a particle detector D is placed in a balloon that travels to a height of 5000 m. Tier-1: Is the number of detected particles at 5000 m (**B**) **higher** than at the start at 0 m height (**A**)? (yes/no) Tier-2: Explanation: (free response)
CM 	**Hiking with a particle detector** Imagine a particle detector D is placed on top of a mountain at 5000 m height. Tier-1: Is the number of detected particles on top of the mountain (**B**) **higher** than in the valley (**A**)? (yes/no) Tier-2: Explanation: (free response)
CO	**Particle detectors on the Earth and the Moon** Imagine a particle detector D is placed on the Moon. Tier-1: Is the number of detected particles on the Moon (**B**) **higher** than on the surface of the Earth (**A**)? (yes/no) Tier-2: Explanation: (free response)
CS 	**Particle detectors on the Earth and on the International Space Station** Imagine a particle detector D is placed outside on the International Space Station. Tier-1: Is the number of detected particles on the Space Station **higher** than on the surface of the Earth? (yes/no) Tier-2: Explanation: (free response)

Table 29 summarizes the results of the categorisation in the pre- as well as the post-test. In total, 93 written explanations distributed over four different items were analysed from the pre-test; 53 were analysed from the post-test. Across all four items, the most frequent explanations were:

- Shielding by the Earth's atmosphere (correct, 33 in the pre-test / 24 post-test)
- The detectors will measure the same number of particles because it is the same atmosphere (wrong, 10 in the pre-test)
- Space is the source of the detected particles / cosmic particles (correct, in total 20 in the pre-test and 24 in the post-test)
- Earth is the only source of the detected particles (wrong, 6 in the pre-test)

Table 29. Results of the qualitative analysis of the free response questions of items CB, CM, CO, CS before and after participating in the S'Cool LAB hands-on session. The table shows the number of coded statements per item and explanation category, the total number of statements per category (n) and the percentage of coded statements.

	Items pre-test						Items post-test					
Category	**CB**	**CM**	**CO**	**CS**	**n**	**%**	**CB**	**CM**	**CO**	**CS**	**n**	**%**
Shielding	9	7	10	7	33	**37%**	4	4	9	7	24	**45%**
Same atmosphere	3	3	2	2	10	**11%**	0	0	0	0	0	**0%**
Space = source	2	1	1	0	4	**4%**	5	2	1	1	9	**17%**
Cosmic particles	4	4	1	7	16	**18%**	7	6	0	2	15	**28%**
Fewer air molecules	4	3	0	3	10	**11%**	0	0	0	0	0	**0%**
Earth = source	3	2	1	0	6	**7%**	0	0	0	0	0	**0%**
Magnetic shielding	0	0	1	0	1	**1%**	0	0	0	0	0	**0%**
Other	1	5	3	4	13	**14%**	1	0	2	2	5	**9%**
#statements	**26**	**25**	**19**	**23**	**93**		**17**	**12**	**12**	**12**	**53**	

Finally, 13 statements in the pre-test and 5 of the post-test were assigned to the 'other' category. For item CB (Balloon), this included statements about the properties of air showers (particles decaying or annihilating on their way to the surface). Students explaining item CM (Mountain) discussed, for example, the reflection of particles by the ground. Others thought that the same ground would lead to the same number of particles, while ignoring the height of the detector. *"Particles are not affected by the height they are in"* describes a similar idea. When discussing particle detectors on the Moon for item CO (Moon), students referred to the different radioactive rocks leading to different detection rates. Furthermore, two students argued that the Earth is bigger which will lead to a higher number of detected particles. Another student explained that there is the same number of particles everywhere. Additional statements mentioned the lack of air showers in space as well as the dependence of the detection rate on the position of the Moon. Explanations of item CS (ISS) include *"particles do not annihilate in space"* or *"air showers are not possible*

in space" associated with a lower expected count rate, and *"radiation comes from all directions"* associated with expectations of no change in the particle count rate. One student explained that the gravitational effect is stronger on Earth, which will cause a higher particle count rate on Earth.

In general, many answers could be interpreted with the expected historic misconception and students' unawareness of cosmic particles. However, it came as a surprise that several students argued with the properties of the atmosphere and the number of air molecules when discussion the count rate of a particle detector. Here, it seemed evident that students do not interpret the term 'particle' as intended. Indeed, in chemistry and physics school education, 'particle' is often used to describe atoms, molecules, or even dust particles. Hence, it is no surprise that students who are unfamiliar with high-energy physics expect 'particle detectors' to detect the 'particles' they know of from school, such as air molecules. This suggests a challenging problem for particle physics education in general.

Based on these results, the items were adapted for the main study. First, tier-1 was changed from a yes-no question to a multiple-choice question with the following three options, because some students did not expect any difference between the two detectors:

- Yes, the particle detector on […] will detect a HIGHER number of particles.
- No, the particle detector on […] will detect a LOWER number of particles.
- No, there won't be a difference.

Furthermore, the following answer options were included in the main study at tier-2 (explanation) to represent typical reasoning patterns:

- The properties of the atmosphere [..] are the same.
- The properties of the ground […] are the same. (not for space items)
- The Earth is the main source of the detected particles.
- Space is the main source of the detected particles.
- The atmosphere has a shielding effect.
- There are fewer air molecules […].

6.4.3 X-ray workshop: irradiation vs. contamination

During X-ray workshops in S'Cool LAB, students work with full-protection X-ray machines to study, for example, the absorption of X-ray photons in matter. Students also plan and conduct an experiment to test the commonly-held hypothesis that food becomes radioactive after exposure to X-radiation. Indeed, previous research results suggest that many students think exposure to radiation will turn stable atoms into radioactive atoms

(see subsection 2.4.2). Table 30 lists four multiple-choice items that were tested with a small student group ($N = 17$) in S'Cool LAB.

Table 30. Item development for X-ray workshop: irradiation vs. contamination. The table lists the item labels, item descriptions and pictures.

Item label, picture & source	Item text
XC Item based on (Eijkelhof, 1990; Molz, 2016)	**Risks for cleaning staff in the radiology department?** In certain industries, it is common to pay employees according to their professional risk. Is such a "risk premium" justified for the cleaning staff of the radiology department of a hospital? 1) Yes, parts of the device are exposed to X-rays and therefore become radioactive themselves. 2) Yes, because of X-ray machines are always a source of radiation, even if they are out of service. 3) No. There is no danger, because the X-ray units are out of service during the cleaning. 4) No, provided that the entire air of the treatment rooms has been exchanged before the cleaning.
XI Item based on (Eijkelhof, 1990)	**Danger for visitors?** After a sports accident, the foot of a patient is often X-rayed to check whether there are bone fractures. Is it dangerous for the patient's family to be in the vicinity of the patient immediately after the X-ray scan? 1) Yes, the foot has been exposed to X-rays and therefore becomes radioactive itself. 2) Yes. The foot accumulates ionizing radiation, and therefore starts to radiate itself. 3) No, because through X-ray machines no radioactive atoms can be transferred. 4) No, if you wait a bit before visiting the patient.
XS Item based on (Molz, 2016; Prather & Harrington, 2001)	**A strawberry in an X-ray machine** A strawberry is X-rayed for 10 minutes inside an X-ray machine. Is it still okay to eat the strawberry afterwards? 1) Yes. To become radioactive itself, the strawberry needs to be exposed to X-rays much longer. 2) Yes. As long as no radioactive atoms are transferred to the strawberry, the strawberry does not emit radiation. 3) No. Ionizing radiation accumulates in the strawberry, therefore it starts to radiate itself. 4) No. The strawberry has been exposed to X-rays and therefore becomes radioactive itself.
XW Item based on (Eijkelhof, 1990)	**A watch in an X-ray machine** A watch is X-rayed for 10 min inside an X-ray machine. Is it still okay to wear the watch afterwards? 1) Yes. To become radioactive itself, the watch needs to be exposed to X-rays much longer. 2) Yes. As long as no radioactive atoms are transferred to the watch, the watch does not emit radiation. 3) No. Ionizing radiation accumulates in the watch, therefore it starts to radiate itself. 4) No. The watch has been exposed to X-rays and therefore becomes radioactive itself.

These items were adapted from Molz (2016) and Prather and Harrington (2001) or were based on results by Eijkelhof (1990). Contrary to the items for the electron tube and cloud chamber workshops, these items already included suitable distractors. Therefore, the pilot was not used to develop these items further but to select two of them for the main study. Table 31 summarises the results of 17 students who answered these four questions before taking part in the S'Cool LAB hands-on session. Although the sample size is very small, these results provide a first estimate of the item easiness and the performance of the distractors.

Item XI (danger of X-ray imaging for visitors of a patient) is very easy for students. Most of them might have had an X-ray scan before or know this situation from family members. Therefore, all of them agree that there is no danger for visitors of patients who had an X-ray scan. Items XS and XW are very similar. However, slightly more students think it is dangerous to eat an X-rayed strawberry than to wear an X-rayed watch.

Table 31. Distribution of 17 student answers across the four different answer options for the four items.

	XC (X-ray cleaning)	**XI (X-ray image)**	**XS (Strawberry)**	**XW (Watch)**
Option 1	3	0	2	2
Option 2	4	0	**6**	**7**
Option 3	**10**	**15**	8	8
Option 4	0	2	1	0
Easiness	**59%**	**88%**	**35%**	**41%**

6.4.4 Selection of concept test items for the main study

Following this test of 13 concept test items with a small group of students, only seven items were selected for the main study to assess selected aspects of students' learning with respect to the three hands-on workshops in S'Cool LAB without causing too much additional test load. Here, Table 32 summarises the proportion of correct answers at the tier-1 levels of the concept test items introduced above. In particular, students had to decide whether or not an electrically charged particle experiences a force in a magnetic field, or whether or not the count rate of a particle detector would be higher in a specific situation. For the X-ray questions, they had to choose the correct answer among four options. Although the small sample size provided only a first estimate of the item properties, the item easiness was used together with other considerations to select the most promising items for the main study.

Item ET (electron beam) was discarded because it was very difficult for the students even after instruction. Indeed, while the other Lorentz force items required students to deduce

the Lorentz force based on the particle's movement as well as the direction of a magnetic field, item ET required more abstract thinking: students had to analyse a curved trajectory to deduce the direction of the magnetic field. This item worked well for university-level students as part of the CSEM and BEMA test instruments (Ding et al., 2006; Maloney et al., 2001), but seemed to be too difficult for high-school students.

Item ED (electron moving downwards through magnetic field) was discarded in favour of the very similar item PD (proton moving downwards through magnetic field). Since students manipulate electrons in S'Cool LAB, it seemed more appropriate to measure their conceptual understanding with a different particle type to avoid testing merely knowledge acquisition. Furthermore, reasoning with the right-hand-rule is usually easier for electrically positively charged particles, such as protons.

Table 32. Overview of properties of conceptual understanding items including sample size per item (n), item easiness (P), corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}) before (pre-test) and after (post-test) the S'Cool LAB hands-on session.

	Pre-test				Post-test			
Description	**n**	**P**	r_{tj}	α_{tj}	**n**	**P**	r_{tj}	α_{tj}
Lorentz force			n=23	n=23			n=33	n=33
(PD) Proton downwards	50	0.30	0.42	0.67	36	0.67	0.62	0.65
(PR) Proton to the right	57	0.32	0.53	0.63	39	0.67	0.54	0.68
(PS) Proton at rest	52	0.62	0.38	0.67	37	0.70	0.23	0.79
(ED) Electron downwards	45	**0.27**	0.51	0.63	38	0.58	0.63	0.64
(ET) Electron deflected	25	**0.28**	0.47	0.65	35	0.43	0.52	0.69
Cronbach's alpha				**0.70**				**0.74**
Cosmic particles			n=31	n=31			n=33	n=33
(CB) Balloon	36	0.44	0.51	0.74	36	**0.92**	0.30	0.65
(CM) Mountain	34	0.47	0.58	0.70	36	0.86	0.35	0.62
(CO) Moon	32	0.63	0.49	0.75	34	0.74	0.52	0.49
(CS) ISS space	32	0.66	0.70	0.64	34	0.85	0.54	0.49
Cronbach's alpha				**0.77**				**0.63**
X-rays			n=17	n=17				
(XC) X-ray cleaning	17	0.59	**0.06**	0.68				
(XI) X-ray image	17	0.88	0.30	0.48				
(XS) X-ray strawberry	17	0.35	0.55	0.23				
(XW) X-ray watch	17	0.41	0.45	0.33				
Cronbach's alpha				**0.53**				

Item CB (particle detector on a balloon flight) was discarded because many tutors tell students about Victor Hess and his balloon flights that resulted in the discovery of cosmic particles and the Physics Nobel Prize in 1936 during the cloud chamber workshop in S'Cool LAB. This was reflected in a very high number of 92% correct answers in the post-

test. In this case, the item did probably not measure conceptual understanding but rather the acquisition of knowledge about this particular case.

Item CO (particle detector on the Moon) was discarded because it required knowledge about the radioactive properties of the Moon's soil in addition to knowledge about the shielding effects of an atmosphere as well as the source of cosmic particles.

In general, all items about cosmic particles showed a significant learning effect from the pre- to the post-test. It seemed that these items measured rather declarative knowledge than conceptual understanding. Here, including a second tier to the item (explanation) might help to increase the item difficulty.

Item XI (X-ray imaging) was discarded for the main study because it was already very easy for students even before taking part in the X-ray workshop in S'Cool LAB. Item XS (strawberry in an X-ray machine) was discarded in favour of the very similar item XW (watch in an X-ray machine). The selected items XC and XW were rephrased slightly to reduce the text and make them easier to understand. Here, item XC was rephrased to *"Is it dangerous for the cleaning staff of a hospital when they clean a room with an X-ray machine?".* Item XW was rephrased to *"A watch is X-rayed for 10 min inside an X-ray machine. Is it still okay to wear the watch afterwards?"*

The final version of these seven items, including the distractors at the different tiers, are included in the pre- and post-visit questionnaires of the main study in appendices 10.3 and 10.5.

6.5 Pre-post effects of the S'Cool LAB hands-on session

To obtain a first estimate of the effect of the S'Cool LAB hands-on session on students' educational outcomes, dependent t-tests were computed to compare the dispositional variables from the pre-visit questionnaire with the situational variables from the post-visit questionnaire, see Table 33. After the hands-on session in S'Cool LAB, students who took part in the pilot study reported significantly higher interest and physics-related self-beliefs with large effect sizes ($d_C = 0.9$ for interest and $d_C = 0.8$ for self-beliefs). They also reported higher self-beliefs related to doing experiments ($d_C = 0.6$, medium-sized effect). However, all effect sizes should be interpreted cautiously as their confidence intervals are very large, see Table 33. Besides, there was no significant change in students' curiosity towards particle physics. Instead, their curiosity remained at a very high level.

Table 33. Pre-post-test differences for students taking part in the pilot study including mean (M), standard deviation (SD), sample size (n), correlation between pre- and post-test scores (r), t-statistic (t), p-value (p), effect size Cohen's d (d_C) and 95% confidence interval.

	M	SD	n	r	t	p	d_C	95% CI
Physics interest								
Dispositional	3.91	1.14	42	0.51	5.82	<0.001	**0.9**	[0.4, 1.3]
Situational	4.84	0.92						
Physics self-beliefs								
Dispositional	3.97	1.09	60	0.33	6.42	<0.001	**0.8**	[0.3, 1.2]
Situational	4.91	0.84						
Experiments self-beliefs								
Dispositional	4.52	0.91	52	0.45	4.56	<0.001	**0.6**	[0.1, 1.0]
Situational	5.05	0.75						
Curiosity state								
Pre-visit (4 items)	4.73	1.01	57	0.74	1.22	0.23		
Post-visit (9 items)	4.58	1.35						

Students also had the chance to leave comments in a free-response field at the end of their post-visit questionnaire. 14 students used this opportunity, mostly to express their gratitude (e.g., *"CERN was the most amazing experience of my life. Thank you so much for everything!"*). The following quote highlights the positive effect on a student's self-beliefs and career aspirations: *"I had the idea of studying physics from before, but everyone kept telling me that it was crazy and it had pretty few professional exits. Thanks to your experiments I am convinced now that I want to at least try it."*

6.6 Summary of the pilot study and conclusions for the main study

Whereas the three workshops forming the S'Cool LAB hands-on session were developed in an iterative design process in 2015, the questionnaires were piloted in summer 2016. Here, the answer of 106 students filling out the pre-visit questionnaire and 60 students filling out the post-questionnaire were analysed to assess the items' and scales' properties. Most of the scales were established by previous studies but had to be adapted to the S'Cool LAB context and tested with an international audience. In general, almost all items showed acceptable properties, and except for the novelty experience scale, all scales showed good internal consistencies in accordance with previous studies.

However, the goal of measuring conceptual understanding in addition to the other educational outcomes required a significant amount of additional questionnaire time. Therefore, the questionnaire was shortened for the main study. Indeed, item and scale analysis of the pilot study data helped to eliminate items without compromising the consistency of the scales. In this way, the 207 items asked in total in the pilot study (64 in

the pre-visit questionnaire and 143 in the post-visit questionnaire) were reduced to 165 items (52 pre / 113 post), see Table 11.

This allowed adding 22 items in both the pre- and post-visit questionnaires to assess students' conceptual understanding and their confidence, which led to 209 items in total for the main study (74 pre / 135 post). As described above, these concept test items were developed with a different small student sample. Although most questions were based on existing items documented by previous studies, they had to be adapted to the sample and the context of S'Cool LAB. Therefore, free-response questions within multi-tier items were used to assess and categorise students' reasoning and to generate suitable distractors for the multiple-choice items of the main study. The seven selected multi-tier items cover key concepts of the three hands-on workshops in S'Cool LAB. Although the small item number limits the scope of these concept items, the results of the pilot study suggested promising item properties for the main study.

7 Main study – evaluating the impact of S'Cool LAB

From October 2016 to June 2017, the main study was conducted to measure the impact of the S'Cool LAB hands-on session on students' educational outcomes and to model these outcomes based on novelty influence and experience factors. The study framework and design are described in section 7.1 and details about the sample are provided in section 7.2. All items and scales were tested in the pilot study, which also informed the selection of the items (see chapter 6). However, due to the small sample size of the pilot study, no factor analyses were performed to validate the dimensionality of certain scales. Hence, section 7.3 describes the items and scale properties including their factor structure. This section also includes the descriptive analysis of a subset of items that were not relevant to answering the research questions but were included in the framework of S'Cool LAB's formative assessment. Subsequently, section 7.4 quantifies the educational effects of the hands-on sessions in S'Cool LAB by comparing student variables that were measured in the pre-visit questionnaires as well as in the post-visit questionnaires. To analyse the intertwined nature of S'Cool LAB's educational outcomes, section 7.5 describes correlations among the eight outcome variables. Section 7.6 answers the additional questions that were part of S'Cool LAB's formative assessment. Then, students' educational outcomes are modelled based on relevant novelty influence and experience factors, both in single-level sequential linear regression models (section 7.7) and in two-level linear models (section 7.8). To summarise the results of the previous two sections, section 7.9 compares the single-level and the two-level models and summarises the results with respect to the overarching research questions.

7.1 Study framework and research design

The main study employed a single group, pre- and post-test research design with educational outcomes as dependent variables, novelty influence and experience factors as independent variables, and age, gender, and group membership as control variables. The 4.5-hours hands-on session in S'Cool LAB with three workshops (electron tube, cloud chamber and X-rays workshop) was developed and tested in an iterative design process prior to this study, see also section 4.5 for a detailed description of the S'Cool LAB workshops. Here, the design process of the workshops was guided by documented students' conceptions, the implementation of predict-observe-explain tasks, and by reducing the extrinsic cognitive load generated by equipment and instructions.

Teachers applied on behalf of their student group to take part in the S'Cool LAB Day programme at CERN, which included a 4.5-hour hands-on session in S'Cool LAB (see section 4.4). Successful teachers then received instructions on how their students can access an online platform that included the pre- and post-visit questionnaires as well as suggestions for preparation material[8] 30 days prior to their trip to CERN.

The different time points of the questionnaires surrounding the S'Cool LAB Day programme are shown in Figure 15. First, students filled out a questionnaire online approximately 30 days prior to their visit to CERN (time t_1). This questionnaire was used to assess students' dispositions and stable character traits. Also, additional novelty influence factors, such as students' previous out-of-school learning experiences were assessed. A sub-sample of 66% of the students filled out a short version of the first questionnaire again onsite directly before starting with the hands-on session in S'Cool LAB (time t_2) to confirm the stability of their dispositional physics interest and self-beliefs as well as their curiosity towards particle physics. Approximately ten days after their trip to CERN, students filled out the post-visit questionnaire (time t_3), which assessed the various educational outcomes and their novelty experience in S'Cool LAB.

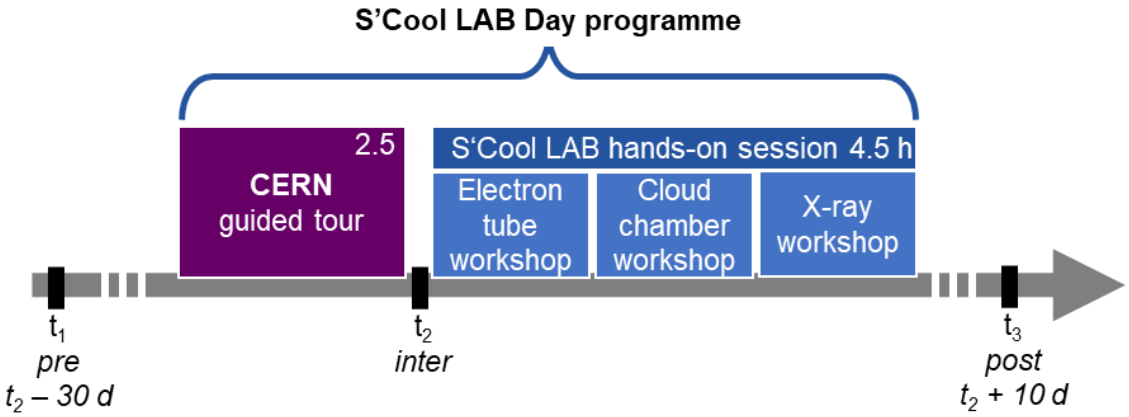


Figure 15. Timeline of the S'Cool LAB Day programme and questionnaires for the main study.

Most of the learning activities in S'Cool LAB are guided using worksheets written in English, which also contain instructions on how to use the high-tech equipment. Workshops are facilitated by volunteering scientists (40% female in the main study), so-called tutors, who help students understand their observations and discuss their findings with them. Although every effort was made to match native-speaking volunteers with the language requirements of the groups, 24% of the workshops took place in English instead of the students' native language.

[8] The design and evaluation of the preparation material was part of a different PhD project and is thus not included in this document.

7.2 Sample

Between October 2016 and June 2017, 29 groups of high-school students from 15 different countries (Austria, Belgium, Czech Republic, France, Germany, Hungary, Italy, Norway, the Netherlands, Poland, Portugal, Singapore, Spain, Switzerland, and the UK) took part in the S'Cool LAB Day programme. 781 students (33% female) aged 15 to 19 ($M = 17.0\,y, SD = 0.9\,y$) filled out the online pre-visit questionnaire before their visit (1-10 days in advance). 534 students filled out the post-visit online questionnaire after their visit (1-10 days after), but only 509 students filled out both the pre-visit and post-visit questionnaires. Subsequently, only students who filled in both the pre- and post-visit questionnaires were considered. The reduced sample consisted of 509 students aged 15 to 19 ($M = 17.0\,y, SD = 0.92\,y$) with 36% female students. Due to a delay in the development of items designated to measure conceptual understanding, only 457 of the 509 students filled out the complete item set. Therefore, analyses related to students' conceptual understanding are based on a slightly smaller sample.

To compare the study sample with other student sample samples, students' favourite and least favourite school subjects were measured and evaluated. Indeed, Zwick and Renn (2000) tried to explore the reason behind the choice of the university studies by interviewing and distributing questionnaires to high-school students (grades 12 & 13, $n = 431$) in the Stuttgart area, Germany in 1999. In their study, the reported favourite subjects and the choice of advanced level school courses were highly related: 86% of the students who favoured a natural science subject had also chosen a corresponding advanced level course. The authors were also interested in the favourite school subjects because they seemed to be influenced strongly by students' interests.

Figure 16 compares the participants in the study by Zwick and Renn (2000) with the sample of the S'Cool LAB main study and shows that S'Cool LAB participants are somewhat special. While physics and mathematics were in the top 4 of students' least favourite school subjects in the study by Zwick and Renn (2000), almost half of the students taking part in the S'Cool LAB main study listed physics and mathematics among their two favourite subjects. In contrast, S'Cool LAB participants disliked Religion, History and Art far more than the students in the study by Zwick and Renn (2000). This effect can be explained by the selection of the student groups and the age range of the students: Many S'Cool LAB participants followed advanced level physics courses and came to CERN in the framework of special physics school trips. Indeed, 64% of the students reported having more than 180 minutes of physics lessons per week and 34% even spent more than 240 minutes on physics per week, see Table 34.

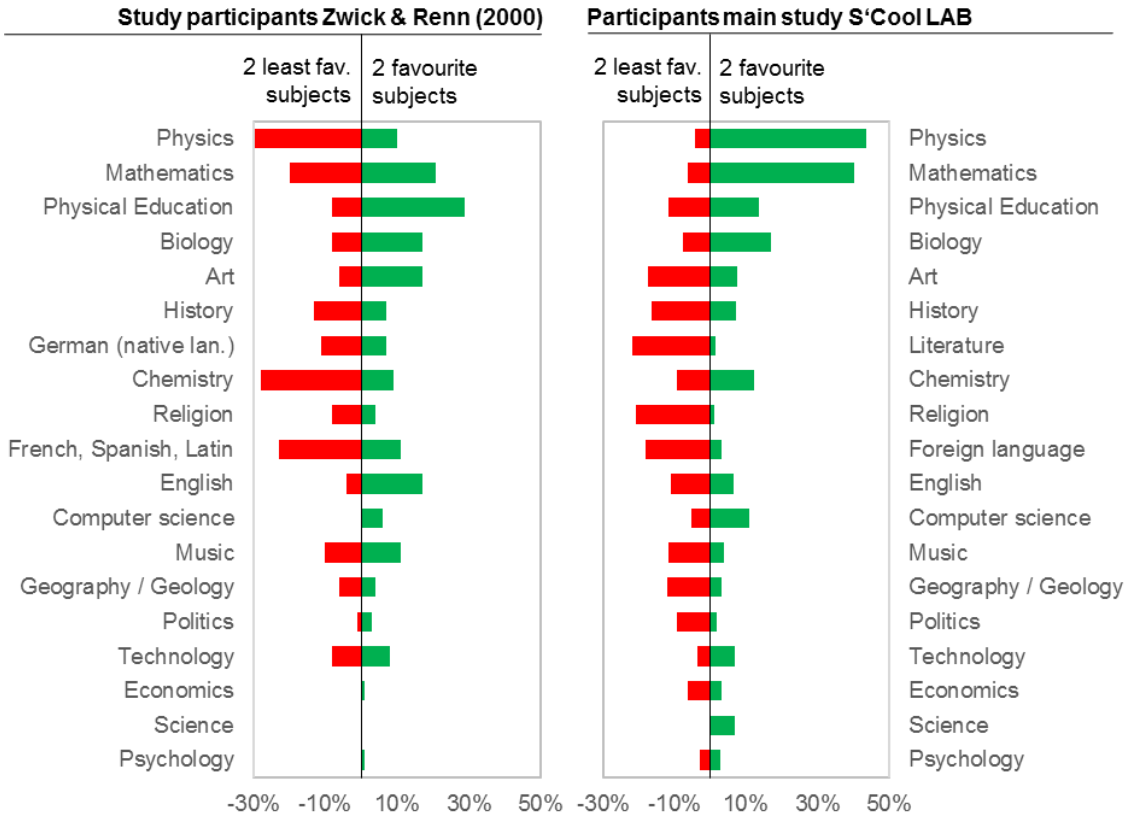


Figure 16. Percentage of students mentioning school subjects among their two favourite subjects (green bars, positive side) or two least favourite subjects (red bars, negative side). On the left: 431 students taking part in a study in Germany 1999 (Zwick & Renn, 2000). On the right: 509 students taking part in the S'Cool LAB main study from 16 different countries in 2016/2017.

The high amount of time spent in physics lessons indicates that many students followed advanced level courses, which is also reflected in the small proportion of only 36% female students. However, the study by Zwick and Renn (2000) based on a German sample is now almost 20 years old, which may limit its comparability with today's students. Moreover, students taking part in the S'Cool LAB Day programme came from different countries and educational systems, which may also contribute to differences in the students' answers.

Table 34. "*How long (in minutes) per week do you spend in physics classes?*"
Free-text answered were collapsed into intervals

Duration	Number of students	% of students
0-120 min	44	9%
121-180 min	141	28%
181-240 min	151	30%
>240 min	172	34%
missing	1	

7.3 Item and instrument analysis

The following section reports the properties of items and scales used during the main study. All items and scales were already tested during the pilot study. Here, section 6.3 includes more details about the origin and theoretical background of the items and their adaptation to the S'Cool LAB context. However, the dimensionality of the scales and their behaviour in a larger sample were analysed only as part of the main study before running statistical inference tests on the scales' mean values. Before examining the properties of items and scales, all items were scanned for coding and data transfer errors by analysing their minimum, mean, and maximum values. Then, principal axis factoring was used to validate the scales' dimensionality as compared to the structure predicted by the underlying theoretical constructs. Consequently, items were sorted into the respective subscales. Some scales were administered before and after the S'Cool LAB hands-on session. Here, only the pre-visit data was analysed through principal axis factoring because this data is generated based on dispositional character traits. Due to the underlying theoretical constructs (see section 2.2), disposition are considered more stable than the situational character states measured after the students' trip to CERN and S'Cool LAB, and hence more suitable for factor analysis.

Consequently, items were grouped into scales while evaluating their properties (such as corrected item-to-total correlation r_{tj}) as well as the scales' properties (such as internal consistency measured by Cronbach's α_C). Provided that no outlying items were identified by this process, scale mean values were calculated and transformed into z-scores. Afterwards, these z-scores were analysed for outliers and compared to a normal distribution. For separate items assessing, for example, students' school grades or previous experience with hands-on experiments, this procedure was similar though without scale-specific analyses. In the following, only short descriptions of the item texts are reported. The full version of all items is included in the original questionnaires of the main study in appendices 10.3 (pre-visit questionnaire), 10.4 (inter-visit questionnaire) and 10.5 (post-visit questionnaire).

7.3.1 Interest and self-beliefs

Since the original basis of both the interest and self-beliefs scales used in this study was a combined motivation test (Hochberg, 2016; Kuhn, 2010; Molz, 2016), these scales with 15 items in total were analysed together. 509 students rated their agreement with 15 statements in the pre-visit questionnaire from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). A KMO value of 0.92 confirmed that the data was suitable for

factor analysis. Inspection of the scree plot after an exploratory principal axis factoring confirmed three factors, see Figure 17, all of which with eigenvalues greater than 1.

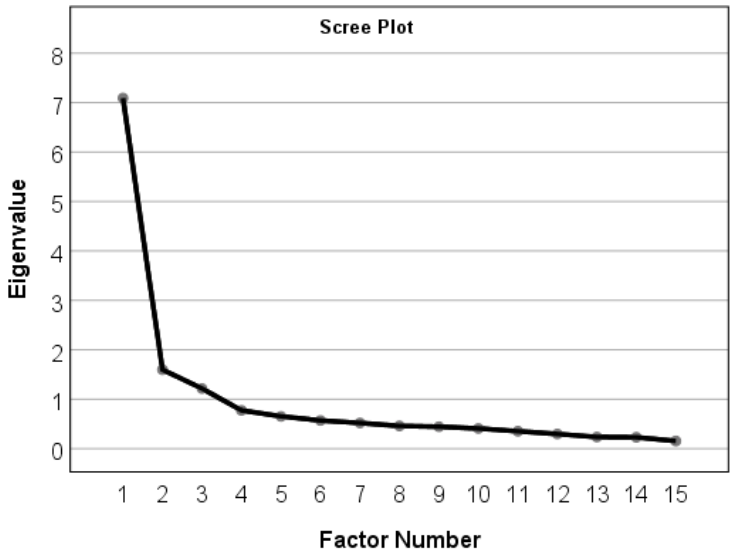


Figure 17. Scree plot of students' motivational dispositions as reported in the pre-visit questionnaire.

Following a Varimax rotation with Kaiser Normalization (converging after 5 iterations), factor loadings ≥ 0.32 are displayed in Table 35 for all 15 items. The three extracted factors can be identified with the theoretical constructs underlying the item selection as follows:

- interest in physics (factor 1, accounts for 24% of the variance after rotation),
- physics self-beliefs (factor 2, accounts for 18% of the variance), and
- experiment self-beliefs (factor 3, accounts for 17% of the variance).

Table 35. Results of principal axis factoring with Varimax rotation & Kaiser Normalization. Factor loadings <0.32 are not shown.

		Factor		
Label	**Short description of the item**	**1**	**2**	**3**
IE1	I spend time on physics in free time	**0.58**	0.32	
IE3	I enjoy solving physics problems	**0.63**	0.39	
IE4	I discuss physics with friends or family	**0.49**		
IE7	I do not notice time passing	**0.56**		
IE8	I like physics classes/S'Cool LAB	**0.71**		
IE9	I want to spend more time on physics/S'Cool LAB	**0.74**		
SC2	I understand the learning material	0.43	**0.51**	
SC3	I look forward to physics classes/ my time in S'Cool LAB	**0.71**	0.33	
SC5	My performance in physics classes/S'Cool LAB is good		**0.82**	
SC7	I find it easy to solve tasks	0.35	**0.74**	
SC10	My classmates think I am good		**0.75**	
SCE2	Doing experiments is easy for me			**0.84**
SCE3	I am good in doing experiments			**0.87**
SCE4	I find it easy to finish in time			**0.64**
SC9	Experiments help me to understand the learning material	0.35		**0.42**
	Modelled variance:	**24%**	**18%**	**17%**

Together, these three factors accounted for 59% of the variance in the variables. Except for item SC3, all items show the highest loadings on the expected factor, although six items have cross-loadings on other factors. As a result, item SC3 was considered to be part of factor 1, namely interest in physics. Here, Hochberg (2016) reports the same behaviour and procedure for this item. Indeed, its item text "*I look forward to physics classes*" seemed to be closely related to the epistemic or feeling-oriented component of interest.

Physics interest

All seven items measuring students' dispositional interest in the pre-visit questionnaire and situational interest in the post-visit questionnaire demonstrated suitable properties leading to an internally consistent scale with $\alpha_C = 0.87$ for both the pre- and post-visit questionnaire, see Table 36. Furthermore, 335 students answered the seven items about their dispositional interest again right before starting the experiments but after taking part in a guided tour at CERN. Also during this inter-questionnaire, the scale demonstrated a high internal consistency with $\alpha_C = 0.90$.

Table 36. Item and scale properties of dispositional and situational interest rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Physics interest		**Pre (n = 509)**				**Post (n = 509)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
IE1	I spend time on physics in free time	3.69	1.37	0.65	0.85	3.51	1.50	0.54	0.88
IE3	I enjoy solving physics problems	4.57	1.12	0.69	0.85	5.05	0.98	0.72	0.84
IE4	I discuss physics with friends or family	4.35	1.34	0.55	0.87	4.91	1.15	0.67	0.85
IE7	I do not notice time passing	4.11	1.27	0.61	0.86	4.84	1.15	0.70	0.84
IE8	I like physics classes/S'Cool LAB	4.74	1.13	0.64	0.86	5.39	0.86	0.68	0.85
IE9	I want to spend more time on physics/S'Cool LAB	4.19	1.33	0.72	0.84	5.16	1.06	0.75	0.84
SC3	I look forward to physics classes/ my time in S'Cool LAB	4.38	1.14	0.72	0.85	5.01	1.02	0.58	0.86
	Cronbach's α_C 95% Confidence interval	**0.87** [0.85, 0.89]				**0.87** [0.85, 0.89]			

The mean values of the scales were z-transformed and compared to a normal distribution. Here, the distributions of the z-values of the interest scales were negatively skewed, both for the dispositional and the situational interest. Indeed, the situational interest scale suffers from a slight ceiling effect. The heavy tails of the distributions include six extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution, these cases were not handled as outliers. Appendix 10.7.2 includes all details about the distributions and extreme values.

Physics self-beliefs

The remaining four items measuring students' dispositional and situational physics self-beliefs provided an internally consistent scale with $\alpha_C = 0.88$ for the pre- and $\alpha_C = 0.83$ for the post-visit questionnaire, see Table 37. Furthermore, 335 students answered the four items about their dispositional physics self-beliefs again right before starting the experiments but after taking part in a guided tour at CERN. Also during this inter-questionnaire, the scale demonstrated a high internal consistency with $\alpha_C = 0.88$.

The mean values of this scale were z-transformed and compared to a normal distribution. Here, the distributions of the z-values of the physics self-beliefs scales were negatively skewed, both for the dispositional and the situational physics self-beliefs. In addition, the situational self-beliefs scale showed a significant positive kurtosis value. The heavy tails of the distributions included six extreme cases with $z < -3.29$. However, due to the deviation from a normal distribution these cases were not handled as outliers. Here, appendix 10.7.6 includes all details about the distributions and extreme values.

Table 37. Item and scale properties of dispositional and situational physics self-beliefs rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Physics self-beliefs		Pre (n = 509)				Post (n = 509)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
SC2	I understand the learning material	4.79	0.97	0.65	0.88	5.04	0.88	0.68	0.77
SC5	My performance in physics classes/S'Cool LAB is good	4.42	1.16	0.80	0.82	4.80	0.90	0.71	0.76
SC7	I find it easy to solve tasks	4.01	1.15	0.80	0.82	4.66	0.90	0.66	0.78
SC10	My classmates think I am good in Physics	4.07	1.34	0.73	0.86	4.31	1.12	0.59	0.82
	Cronbach's α_C	**0.88**				**0.83**			
	95% Confidence interval	[0.86, 0.90]				[0.80, 0.85]			

Experiment self-beliefs

Furthermore, 472 students[9] answered four additional items about their dispositional self-beliefs related to doing experiments, and 509 students answered these four items after the hands-on session in S'Cool LAB to assess their situational self-beliefs related to doing experiments in S'Cool LAB. The four items formed an internally consistent scale with $\alpha_C = 0.82$ for the pre- and $\alpha_C = 0.77$ for the post-visit questionnaire, see Table 38. Furthermore, 330 students answered the four items about their dispositional experiment self-beliefs again right before starting the experiments but after taking part in a guided tour

[9] By accident, this set of four items was not set 'mandatory' in the first iterations of the online questionnaire, which explains the missing data points for this scale.

at CERN. Also during this inter-questionnaire, the scale demonstrated a high internal consistency with $\alpha_C = 0.83$.

Table 38. Item and scale properties of dispositional and situational experiment self-beliefs rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Experiment self-beliefs		Pre (n = 472)				Post (n = 509)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
SC9	Experiments help me to understand the learning material	4.30	1.12	0.76	0.73	4.68	0.95	0.68	0.66
SCE2	Doing experiments is easy for me	4.35	1.09	0.79	0.71	4.72	0.96	0.73	0.63
SCE3	I am good in doing experiments	4.22	1.14	0.62	0.79	4.51	1.17	0.41	0.81
SCE4	I find it easy to finish in time	4.87	1.09	0.45	0.86	5.06	0.95	0.52	0.74
	Cronbach's α_C	**0.82**				**0.77**			
	95% Confidence interval	[0.80, 0.85]				[0.73, 0.80]			

The mean values of these scales were z-transformed and compared to a normal distribution. But the distributions of the z-values of the experiment self-beliefs scales deviated significantly from a normal distribution due to significant negative skewness and positive kurtosis both for the dispositional and the situational experiment self-beliefs. Also, the heavy tails of the distributions included eight extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution, these cases were not handled as outliers. Here, appendix 10.7.7 lists all details about the distributions and extreme values.

Correlations between motivational variables

As expected, the motivational variables 'physics interest', 'physics self-beliefs' and 'experiment-related self-beliefs' were correlated, see Table 39.

Table 39. Correlations between motivational variables in the pre, inter, and post questionnaires. High correlation ($r > 0.5$) are highlighted in blue.

Correlation between	Questionnaire	Variables	n	r		p
Physics interest and physics self-beliefs	Pre t_1	IE1, SC1	509	**0.65**	***	3E-63
	Inter t_2	IE2, SC2	333	**0.64**	***	7E-40
	Post t_3	IE3, SC3	509	**0.54**	***	2E-40
Physics interest and experiment self-beliefs	Pre t_1	IE1, SCE1	472	**0.53**	***	5E-35
	Inter t_2	IE2, SCE2	328	0.41	***	8E-15
	Post t_3	IE3, SCE3	509	0.49	***	1E-31
Physics self-beliefs and experiment self-beliefs	Pre t_1	SC1, SCE1	472	**0.50**	***	4E-31
	Inter t_2	SC2, SCE2	328	0.42	***	8E-16
	Post t_3	SC3, SCE3	509	**0.81**	***	4E-119

In particular, physics interest and physics self-beliefs showed high and highly significant correlations across all three questionnaires ($r = 0.65|0.64|0.54$***). Furthermore, physics self-beliefs and experiment self-beliefs demonstrated a high correlation in the pre-visit questionnaire ($r = 0.50$***), and an extremely high correlation in the post-visit

questionnaire ($r = 0.81$***). This is no surprise, as the questions assessing students' situational physics self-beliefs focus on students' perceived success with the experimental activities in S'Cool LAB. Hence, these questions have a strong focus on experiments due to the design of the questionnaire.

Consequently, also due to the weaker scale properties, experiment self-beliefs are no longer considered a separate educational outcome in the following. Also, due to missing data in this scale in the pre-visit questionnaire, dispositional experiment self-beliefs were not used as novelty influence factor in the linear models of the educational outcomes. Instead, they were analysed as a separate informative variable.

7.3.2 Curiosity trait and state

Although curiosity trait and state are separate constructs, they have been found to correlate (Litman et al., 2005; Reio Jr et al., 2006). Hence, both scales are analysed together in the following. 509 students rated their agreement with eight curiosity items in the pre-visit questionnaire and six items in the post-visit questionnaire from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). A KMO value of 0.88 confirmed that the pre-visit data is suitable for factor analysis. Inspection of the scree plot after an exploratory principal axis factoring confirmed two factors, see Figure 18, both with eigenvalues greater than 1.

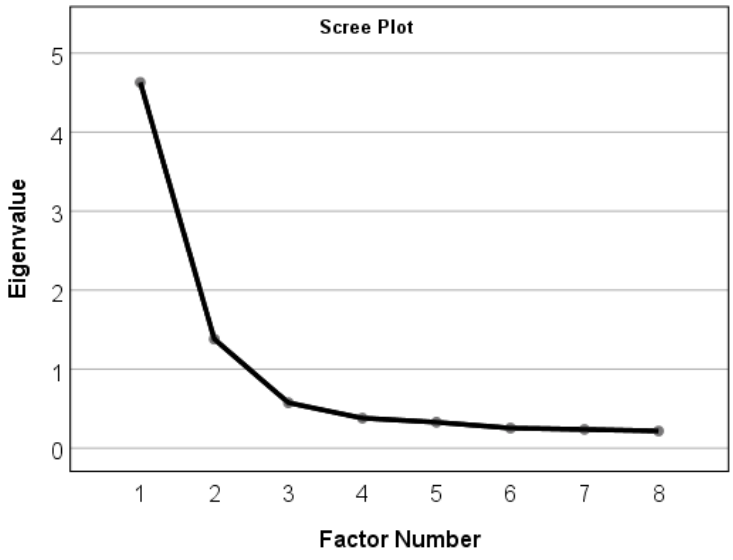


Figure 18. Scree plot of students' curiosity as reported in the pre-visit questionnaire.

Following a Varimax rotation with Kaiser Normalization (converging after 10 iterations), factor loadings ≥ 0.32 are displayed in Table 40 for all eight items. The two extracted factors can be identified with the expected theoretical constructs underlying the item selection as follows:

- Curiosity trait (factor 1, accounts for 39% of the variance after rotation),
- Curiosity state towards particle physics (factor 2, accounts for 29% of the variance).

Together, both factors accounted for 68% of the variance in the variables. All items showed high loadings on the expected factor without cross-loadings.

Table 40. Results of principal axis factoring with Varimax rotation & Kaiser Normalization. Factor loadings <0.32 are not shown.

		Factor	
Label	**Short description of the item**	**1**	**2**
CT1	I find it fascinating to learn new things	0.79	
CT2	I am curious about new things	0.84	
CT4	I like to learn things that are new to me	0.84	
CT5	I find new things inspiring	0.72	
CT6	When I hear something new, I want to know more	0.61	
CS2	I want to learn more about particle physics		0.87
CS8	I want to understand the topic better		0.81
CS9	I find it fascinating to spend time on particle physics		0.75
	Modelled variance:	**39%**	**29%**

Curiosity trait

Although all five items showed acceptable properties and the scale displayed good internal consistency with $\alpha_C = 0.89$, the items suffer from a ceiling effect with mean values well above 5 ($\cong 80\%$ of the scale maximum), see Table 41. This is also reflected in the severe deviation from a normal distribution of the z-transformed mean values with a high negative skewness value and a high positive kurtosis value, as illustrated in appendix 10.7.8. The heavy tail of the distribution included four extreme cases with $z < -3.29$. However, due to the deviation from a normal distribution, these cases were not handled as outliers.

Table 41. Item and scale properties of curiosity trait rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Curiosity trait		**Pre (N = 509)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CT1	I find it fascinating to learn new things	5.44	0.77	0.76	0.87
CT2	I am curious about new things	5.39	0.78	0.79	0.86
CT4	I like to learn things that are new to me	5.35	0.76	0.80	0.86
CT5	I find new things inspiring	5.07	0.90	0.72	0.88
CT6	When I hear something new, I want to know more	5.03	0.87	0.65	0.89
	Cronbach's α_C	**0.89**			
	95% Confidence interval	[0.88, 0.91]			

Previous studies employing a similar scale also reported high values of 78% (Hirth, 2019) or 70% of the scale's maximum (Hochberg, 2016). However, students taking part in the

S'Cool LAB main study were not just more interested in physics and mathematics than other students, they were also generally extremely curious with an average curiosity trait value of 85% of the scale's maximum. Hence, due to the saturation of the scale, curiosity trait may not be a suitable predictor of educational outcomes.

Curiosity state towards particle physics

All items measuring students' curiosity state had suitable properties leading to an internally consistent scale with $\alpha_C = 0.88$ in the pre-visit questionnaire (3 items) and $\alpha_C = 0.95$ in the post-visit questionnaire (6 items), see Table 42. Furthermore, 337 students answered the three items of the pre-visit questionnaire about their curiosity towards particle physics again right before starting the experiments but after taking part in a guided tour at CERN. Also during this inter-questionnaire, the scale demonstrated a high internal consistency with $\alpha_C = 0.89$.

The mean values of these scales were z-transformed and compared to a normal distribution. However, the distributions deviated significantly from a normal distribution due to significant negative skewness and positive kurtosis for all three questionnaires (pre, inter, post). Also, the heavy tails of the distributions included ten extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution, these cases were not handled as outliers. Here, appendix 10.7.9 lists details about the distributions and extreme values.

Table 42. Item and scale properties of curiosity state rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Curiosity state		**Pre (n = 509)**				**Post (n = 509)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
CS1	My time in S'Cool LAB drew my attention drawn to the topic					4.83	1.14	0.82	0.94
CS2	I want to learn more about particle physics	5.03	1.02	0.81	0.80	4.89	1.14	0.87	0.94
CS5	S'Cool LAB made me curious to learn more about particle physics					4.86	1.10	0.91	0.93
CS7	I am more curious now					4.88	1.12	0.84	0.94
CS8	I want to understand the topic better	5.06	0.98	0.77	0.84	4.96	1.06	0.86	0.94
CS9	I find it fascinating to spend time on particle physics	4.55	1.15	0.75	0.87	4.79	1.12	0.79	0.95
	Cronbach's α_C	**0.88**				**0.95**			
	95% Confidence interval	[0.86, 0.90]				[0.94, 0.96]			

7.3.3 Test-retest reliability of students' motivational variables and curiosity

Dispositions are stable character traits. Hence, students' dispositional physics interest and self-beliefs should not change between the pre-visit questionnaire, filled out approximately 30 days before their trip to CERN, and the inter questionnaires, filled out at CERN directly after the guided tour but before starting the hands-on session in S'Cool LAB. To provide an estimate of the scales test-retest reliability, 66% of the students taking part in the main study filled out a very short questionnaire with selected items of the pre-visit questionnaire. In addition to the dispositional physics interest and self-beliefs, they also answered again three items assessing their curiosity state towards particle physics. Table 43 summarises the means values of the respective scales for these two time points and the results of the dependent samples t-tests.

Table 43. Stability of students' dispositional interest, self-beliefs and curiosity state including mean (M), standard deviation (SD), sample size (n), correlation between pre- and inter-test scores (r), t-statistic (t) and p-value (p).

		M	SD	n	r	t	p
Dispositional physics interest	IE1	4.27	0.97	335	0.78	2.77	**0.006**
	IE2	4.17	0.98				
Dispositional physics self-beliefs	SC1	4.30	1.02	335	0.80	1.17	0.242
	SC2	4.26	0.97				
Curiosity state particle physics	CS1	4.87	0.96	337	0.69	-3.74	**0.0002**
	CS2	5.01	0.87				

Indeed, students' physics self-beliefs did not differ between the two time points. However, there is a small negative effect with respect to students' dispositional interest ($d_C = 0.2\ (95\%\ CI\ [0.0, 0.3]$) and a small positive effect with respect to students' curiosity state towards particle physics ($d_C = 0.2, 95\%\ CI\ [0.0, 0.3]$). One possible explanation of the slight increase in students' curiosity state could be the effect of the guided tour at CERN. When filling out the inter questionnaire, they had just spent 2.5 hours with a CERN physicist showing them an old particle accelerator and the control room of the largest particle detector at CERN. Also, while students filled out the pre-visit questionnaire at their school, they filled out the inter questionnaire at CERN, which may contribute to some kind of a 'CERN-halo' effect on this variable. In turn, such a CERN-effect favouring students' curiosity towards particle physics could explain a small detrimental effect on students' interest in school physics. Indeed, items assessing students' dispositional physics interest focused on their physics lessons in school, which may not regularly include exciting modern physics topics.

However, these effects are very small, and the 95% confidence intervals include a zero effect $d_C = 0.0$. Also, as a rule of the thumb, Glen (2016) suggest a Pearson correlation $r \geq 0.7$ as an indicator for an acceptable test-retest reliability, which is satisfied by the dispositional variables, and almost satisfied by the curiosity state scale, which displayed a marginally smaller correlation of $r = 0.69$.

7.3.4 Enjoyment, Easiness, Helpfulness

509 students rated their agreement with 15 statements in the post-visit questionnaire from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). A KMO value of 0.90 confirmed that the data was suitable for factor analysis to validate the factor structure based on the results by Rennie (1994). Inspection of the scree plot after an exploratory principal axis factoring confirmed three factors, see Figure 19, all with eigenvalues > 1.

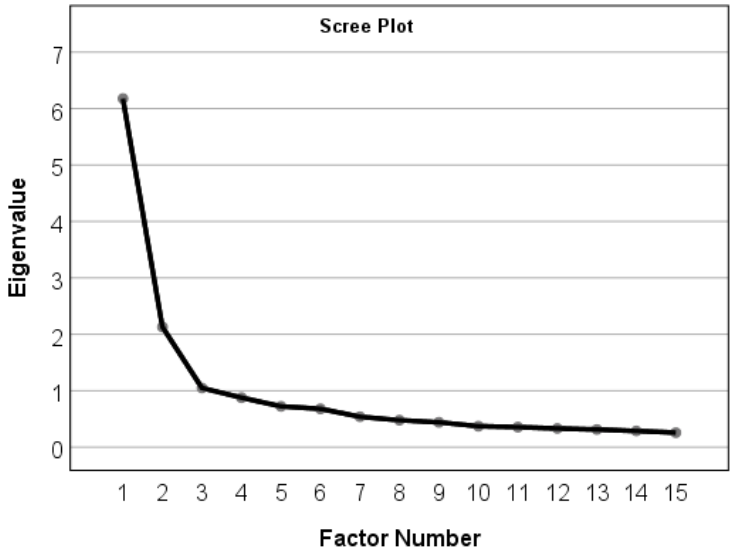


Figure 19. Scree plot of students' perceived affective experience as reported in the post-visit questionnaire.

Following a Varimax rotation with Kaiser Normalization (converging after 8 iterations), factor loadings ≥ 0.32 are displayed in Table 44 for all 15 items. The three extracted factors can be identified with the expected theoretical constructs underlying the item selection as follows:

- helpfulness (factor 1, accounts for 19% of the variance after rotation),
- easiness (factor 2, accounts for 18% of the variance) and
- enjoyment (factor 3, accounts for 16% of the variance).

Together, all three factors accounted for 53% of the variance in the variables. All items showed high loadings on the expected factor. However, three items of the enjoyment scale showed cross-loadings on factor 1, namely helpfulness. Since both sub-scales are related

to the construct interest (helpfulness as part of the value-oriented 'hold' component and enjoyment as part of the feeling-oriented component), it is no surprise that these factors were not exactly orthogonal. Indeed, rotating factors using the oblique rotation method 'Promax' after factor extraction removed these cross-loadings by allowing the extracted factors to correlate, see Table 44.

Table 44. Results of principal axis factoring with Varimax rotation and Promax rotation, both with Kaiser Normalization. Factor loadings <0.32 are not shown.

		Varimax Factor			Promax Factor		
Label	**Short description of the item**	**1**	**2**	**3**	**1**	**2**	**3**
EN1	I enjoyed using the equipment.			0.70			0.83
EN2	I enjoyed preparing the experiments.			0.66			0.76
EN3	I enjoyed doing the experiments.	0.33		0.65			0.73
EN4	I enjoyed discussing the phenomena.	0.37		0.46			0.43
EN5	I enjoyed the time in S'Cool LAB.	0.42		0.67			0.72
EA1	It was easy to understand the instructions.		0.68			0.70	
EA2	It was easy to prepare the experiments.		0.78			0.81	
EA3	It was easy to use the equipment.		0.79			0.83	
EA4	It was easy to get a result.		0.56			0.58	
EA5	It was easy to understand what it is about		0.52			0.53	
HE1	It was helpful for my physics classes	0.65			0.65		
HE2	It helped me to get better at experiments	0.75			0.84		
HE3	It helped me to understand physics	0.71			0.73		
HE4	It helped me get idea how scientists work	0.63			0.66		
HE5	It helped me get idea what CERN does	0.56			0.57		
	Modelled variance:	**19%**	**18%**	**16%**			

Enjoyment

All five items showed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.86$, see Table 45. The mean values of the scale were z-transformed and compared to a normal distribution. Here, the distribution of the z-values was negatively skewed and also showed a significant positive kurtosis value. Indeed, the enjoyment scale suffered from a ceiling effect with mean values well above 5 ($\cong 80\%$ of the scale maximum). The heavy tail of the distribution included three extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution, these cases were not handled as outliers. Appendix 10.7.13 lists more details about the distribution and extreme values.

Table 45. Item and scale properties of perceived enjoyment rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Enjoyment		Post (N = 509)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}
EN1	I enjoyed using the equipment.	5.47	0.71	0.71	0.83
EN2	I enjoyed preparing the experiments.	5.18	0.85	0.68	0.83
EN3	I enjoyed doing the experiments.	5.51	0.74	0.70	0.83
EN4	I enjoyed discussing the phenomena.	5.09	0.97	0.59	0.86
EN5	I enjoyed the time in S'Cool LAB.	5.36	0.82	0.75	0.81
	Cronbach's α_C	**0.86**			
	95% Confidence interval	[0.84, 0.88]			

Helpfulness

All five items showed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.84$, see Table 46. The mean values of the scale were z-transformed and compared to a normal distribution. Here, the distribution of the z-values was negatively skewed. The heavy tail of the distribution included three extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution these cases were not handled as outliers. Appendix 10.7.13 lists details about the distribution and extreme values.

Table 46. Item and scale properties of perceived helpfulness rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Helpfulness		Post (N = 509)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}
HE1	It was helpful for my physics classes	4.84	1.10	0.63	0.80
HE2	It helped me to get better at experiments	4.73	1.14	0.68	0.79
HE3	It helped me to understand physics	5.09	0.92	0.69	0.79
HE4	It helped me get idea how scientists work	4.92	1.13	0.63	0.81
HE5	It helped me get idea what CERN does	5.06	1.06	0.57	0.82
	Cronbach's α_C	**0.84**			
	95% Confidence interval	[0.81, 0.86]			

Easiness

All five items demonstrated acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.81$, see Table 47. The mean values of the scale were z-transformed and compared to a normal distribution. Here, the distribution of the z-values was significantly negatively skewed. The heavy tail of the distribution included two extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution

these cases were not handled as outliers. Appendix 10.7.13 lists details about the distribution and extreme values.

Table 47. Item and scale properties of perceived easiness rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Easiness		**Post (N = 509)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
EA1	It was easy to understand the instructions.	5.13	0.83	0.62	0.77
EA2	It was easy to prepare the experiments.	5.18	0.75	0.68	0.75
EA3	It was easy to use the equipment.	5.20	0.74	0.71	0.75
EA4	It was easy to get a result.	4.92	0.88	0.53	0.80
EA5	It was easy to understand what it is about	4.72	0.96	0.50	0.81
	Cronbach's α_C	**0.81**			
	95% Confidence interval	[0.78, 0.84]			

Summary of enjoyment, easiness and helpfulness scales

The three subscales enjoyment, easiness and helpfulness formed separate constructs as demonstrated by Rennie (1994). However, as detailed above, the S'Cool LAB hands-on session produced exceptionally high values for these affective outcomes, which almost saturated the scales. Nevertheless, they were modelled as separate educational outcomes.

7.3.5 Novelty

509 students rated their agreement with 14 statements about their perception of novelty in the post-visit questionnaire from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). A KMO value of 0.86 confirmed that the data is suitable for factor analysis to compare the factor structure with the results of Molz (2016). Inspection of the scree plot after an exploratory principal axis factoring confirmed three factors, see Figure 20, all with eigenvalues greater than 1.

Following a Varimax rotation with Kaiser Normalization (converging after 5 iterations), factor loadings ≥ 0.32 are displayed in Table 48 for all 14 items. Because two of the items did not clearly load on one of the three factors, an oblique factor rotation was conducted in addition. Indeed, it was expected that the different factors of novelty reception were not strictly orthogonal but instead correlated. Here, the results of a Promax rotation clearly assigned the 14 items to three factors, see Table 48.

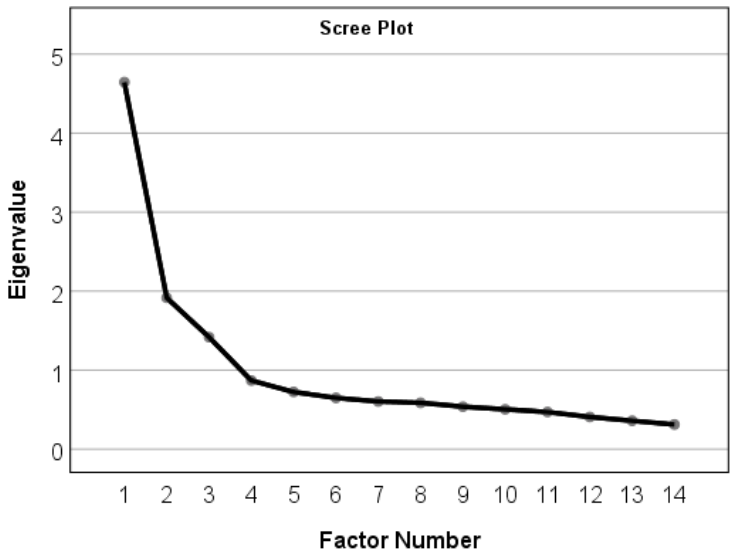


Figure 20. Scree plot of students' perceived novelty as reported in the post-visit questionnaire.

The three extracted factors accounted for 46% of the variance and can be identified as

- tinkering support due to organisation (factor 1, accounts for 21% of the variance after Varimax rotation),
- lack of problems (factor 2, accounts for 13% of the variance) and
- setting orientation due to preparation (factor 3, accounts for 12% of the variance).

Table 48. Results of principal axis factoring with Varimax rotation and Promax rotation, both with Kaiser Normalization. Factor loadings <0.32 are not shown. In addition, the last column includes the identified factor for the item analysed by Molz (2016) with the factors setting orientation (SO), capabilities (Cap), Fear and Fun. Some items have been inverted for this study, which is indicated by a minus sign.

		Varimax Factor			Promax Factor			Molz
Label	**Short description of the item**	**1**	**2**	**3**	**1**	**2**	**3**	**2016**
TOP1	I was familiar with the schedule	0.34		0.41			0.36	SO
TOP2	The time was well organized	0.55			0.58			SO
TOP3	I was well prepared	0.35		0.56			0.53	SO
TOP4	I knew what to expect			0.72			0.78	(SO)
TOP5	I was easily getting familiar with the different work places	0.39		0.40			0.33	-
TOP6	Before arriving at CERN, I knew how I was going to work			0.57			0.62	SO
TOP7	The materials were easy to find	0.42			0.34			-
TIN1-	It was hard to do experiments		0.72			0.73		-Fun
TIN2-	I was afraid to make mistakes		0.64			0.71		Fear
TIN3-	It took time until I was familiar		0.83			0.87		-Cap
TIN4	It was fun to try the experiments	0.61			0.66			Fun
TIN5	I received enough information	0.73			0.77			SO
TIN6	Concentrate on the experiments, without "fighting" the equipment	0.67			0.73			Cap
TIN7	The instructions were helpful	0.74			0.84			-
	Modelled variance:	**21%**	**13%**	**12%**				

Comparison of the extracted factor structure with previous studies

For comparison, Molz (2016) reported four factors when measuring students' perception of novelty after receiving targeted preparation with very similar items, which he interpreted as novelty factor 'setting orientation', novelty factor 'capabilities', 'fun', and 'lack of fear', see also subsection 6.3.4. However, the item texts differed between the study by Molz (2016) and this study. While Molz (2016) introduced all items with "*Thanks to the preparation* …" because his study participants received specific preparation to reduce, for example, novelty associated with setting orientation, the items in this study did not include a reference to specific preparation material.

Factor 1, 'tinkering support due to organisation' includes six items, which were assigned to the factors 'setting orientation', 'capabilities' or 'fun' in the study by Molz (2016). Here, the items of this scale refer to rather organisational features of the out-of-school learning experience that supported students while tinkering with the equipment, such as ease in finding material, helpful, and sufficient instructions.

Factor 2, 'lack of problems' consists of three items, out of which two items had been inverted for the S'Cool LAB study. It shares similarities and one item with the subscale 'fear' by Molz (2016).

Factor 3, 'setting orientation due to preparation' consists of five items reflecting students' feeling of preparedness and ease of orientation. Students received some preparation material prior to their trip to CERN. Hence, this factor represents the factor setting orientation as described previously by Cors et al. (2017) and measured by Molz (2016). Indeed, most of the items in this subscale were assigned to the factor 'setting orientation' by Molz (2016). Item TOP4 was intended to load on the factor 'setting orientation' by Molz (2016) but was excluded from his further analyses since it did not load as expected.

In summary, the three extracted factors can be interpreted in accordance with previous results by Molz (2016) and Cors et al. (2017).

Tinkering support due to organisation

All six items showed acceptable properties, and the scale displayed good internal consistency with $\alpha_C = 0.81$, see Table 49. The mean values of the scale were z-transformed and compared to a normal distribution. Here, the distribution of the z-values was negatively skewed and also showed a significant positive kurtosis value. Indeed, this scale suffered from a ceiling effect with mean values well above 5 ($\cong 80\%$ of the scale maximum). The heavy tail of the distribution included six extreme cases with z-values

below -3.29. However, due to the deviation from a normal distribution these cases were not handled as outliers. Appendix 10.7.17 lists details about the distribution and extreme values. In general, students taking part in S'Cool LAB hands-on sessions felt that the activities were very well organised and they felt very well supported while doing experiments.

Table 49. Item and scale properties of perceived tinkering support rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Tinkering support due to organisation		Post (N = 509)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}
TOP2	The time was well organized	5.19	0.86	0.51	0.80
TOP7	The materials were easy to find	5.28	0.88	0.45	0.81
TIN4	It was fun to try out the experiments	5.42	0.77	0.56	0.79
TIN5	I received enough information	5.28	0.82	0.69	0.76
TIN6	I was able to concentrate on the experiments, without "fighting" with the equipment	5.24	0.86	0.60	0.78
TIN7	The instructions were helpful	5.35	0.76	0.65	0.77
	Cronbach's α_C	**0.81**			
	95% Confidence interval	[0.79, 0.84]			

Lack of problems

All three items showed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.78$, see Table 50. The mean values of the scale were z-transformed and compared to a normal distribution. In particular, the distribution of the z-values was negatively skewed but did not include any extreme cases. Appendix 10.7.17 lists details about the distribution.

Table 50. Item and scale properties of perceived problems rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Lack of problems		Post (N = 509)			
Label	Short description of the item	M	SD	r_{tj}	α_{tj}
TIN1-	It was hard to do the experiments (inversely coded)	4.74	1.18	0.63	0.70
TIN2-	I was afraid to make mistakes (inversely coded)	3.96	1.46	0.56	0.79
TIN3-	It took a long time until I was familiar (inversely coded)	4.61	1.25	0.69	0.63
	Cronbach's α_C	**0.78**			
	95% Confidence interval	[0.75, 0.81]			

Setting orientation due to preparation

All five items demonstrated acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.70$, see Table 51. The mean values of the scale were z-

transformed and did not deviate from a normal distribution. Appendix 10.7.17 lists details about the distribution.

Table 51. Item and scale properties of perceived setting orientation rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Setting orientation due to preparation		Post (N = 509)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
TOP1	I was familiar with the schedule	4.75	1.08	0.41	0.68
TOP3	I was well prepared	4.91	0.98	0.54	0.63
TOP4	I knew what to expect	4.27	1.16	0.55	0.61
TOP5	I was easily getting familiar with the different work places	5.06	0.87	0.42	0.67
TOP6	Before arriving at CERN, I knew how I was going to work	4.13	1.41	0.44	0.68
	Cronbach's α_C 95% Confidence interval	**0.70** [0.66, 0.74]			

7.3.6 Cognitive load

In total, students rated their agreement with 17 statements about their perceived cognitive load in the post-visit questionnaire from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). A KMO value of 0.90 confirmed that the data is suitable for factor analysis. Inspection of the scree plot after an exploratory principal axis factoring suggested four factors, see Figure 21, all with eigenvalues greater than 1.

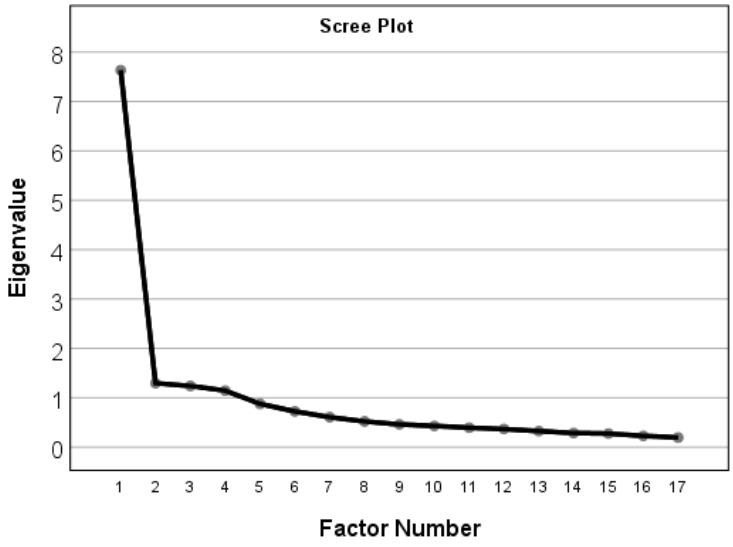


Figure 21. Scree plot of students' perceived cognitive load as reported in the post-visit questionnaire.

Following a Varimax rotation with Kaiser Normalization (converging after 6 iterations), factor loadings ≥ 0.32 are displayed in Table 52 for all 17 items. Together, the four extracted factors accounted for 57% of the variance in the variables and confirmed, overall,

the expected factor structure, although some items showed cross-loadings on other factors:

- cognitive load in S'Cool LAB in general (factor 1, accounts for 17% of the variance after Varimax rotation),
- cognitive load of the electron acceleration workshop (factor 2, 14% of the variance),
- cognitive load of the X-ray workshop (factor 3, 13% of the variance) and
- cognitive load of the cloud chamber workshop (factor 4, 13% of the variance).

Table 52. Results of principal axis factoring with Varimax rotation & Kaiser Normalization. Factor loadings <0.32 are not shown.

		Factor			
Label	**Short description of the item**	**1**	**2**	**3**	**4**
CLE1	I had problems interpreting my observations	**0.60**			
CLE2	The experiments were difficult	**0.59**			
CLE3	I had problems understanding the principles	**0.68**			
CLE4	It was hard for me to understand the instructions	**0.57**			0.33
CLE5	I would have needed more assistance	**0.59**			
CLC1	It was difficult to use the equipment				**0.68**
CLC4	I was overwhelmed by the features				**0.71**
CLC5	I had problems to make observations				**0.57**
CLC6	I had problems understanding the physics	0.42		0.33	**0.27**
CLEL1	It was difficult to use the equipment		**0.73**		
CLEL4	I was overwhelmed by the features		**0.71**		0.41
CLEL5	I had problems to make observations	0.33	**0.68**		
CLEL6	I had problems understanding the physics	0.42	**0.52**		
CLX1	It was difficult to use the equipment			**0.72**	
CLX4	I was overwhelmed by the features		0.40	**0.55**	0.38
CLX5	I had problems to make observations			**0.66**	
CLX6	I had problems understanding the physics	0.45		**0.60**	
	Modelled variance:	**17%**	**14%**	**13%**	**13%**

Indeed, it is somewhat remarkable that the factor analysis was able to separate the perceived cognitive load for the different workshops. This is only possible if students clearly remembered these three different workshops even ten days after their S'Cool LAB visit, and were still able to differentiate their retrospective perception of cognitive load. Thus, smaller cross-loadings for some of the items are not alarming. Only one item (CLC6 *"I had problems understanding the physics principles of the [cloud chamber] experiment."*) did not clearly load on the expected factor. Maybe the extraneous cognitive load of the cloud chamber workshop (covered by the other items in the subscale) was very small compared to the germane load provided by the underlying physics principles (covered by the problematic item CLC6). In the end, only the scale 'cognitive load in S'Cool LAB in general' was considered as predictor when modelling S'Cool LAB's educational outcomes. The

subscales representing the cognitive load related to the three different workshops were analysed separately in the framework of S'Cool LAB's formative assessment.

Cognitive load in S'Cool LAB in general

All five items showed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.85$, see Table 53. The mean values of the scale were z-transformed and their distribution was compared to a normal distribution. Here, the distribution of the z-values was significantly positively skewed and also showed a significant positive kurtosis value. Indeed, the cognitive load scale tended towards a floor effect. The heavy tail of the distribution included three extreme cases with z-values above $+3.29$. However, due to the deviation from a normal distribution, these cases were not handled as outliers. Appendix 10.7.18 lists details of the distribution and extreme values.

Table 53. Item and scale properties of perceived cognitive load rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

		Post (N = 509)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CLE1	I had problems interpreting my observations	2.85	1.21	0.60	0.83
CLE2	The experiments were difficult	2.25	1.04	0.69	0.81
CLE3	I had problems understanding the principles	2.56	1.16	0.67	0.81
CLE4	It was hard for me to understand the instructions	2.00	1.06	0.68	0.81
CLE5	I would have needed more assistance	2.25	1.18	0.65	0.82
	Cronbach's α_C	**0.85**			
	95% Confidence interval	[0.83, 0.87]			

Cognitive load cloud chamber

All four items showed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.78$, see Table 54.

Table 54. Item and scale properties of perceived cognitive load during the cloud chamber workshop rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

		Post (N = 503)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CLC1	It was difficult to use the equipment	1.73	0.91	0.65	0.69
CLC4	I was overwhelmed by the features	2.01	1.13	0.59	0.72
CLC5	I had problems to make observation	1.85	1.02	0.63	0.70
CLC6	I had problems understanding the physics	2.18	1.14	0.47	0.78
	Average	**1.94**	**1.05**		
	Cronbach's α_C	**0.78**			
	95% Confidence interval	[0.74, 0.81]			

Cognitive load electron tube workshop

All four items showed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.85$, see Table 55.

Table 55. Item and scale properties of perceived cognitive load during the electron tube workshop rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

		Post (N = 503)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CLEL1	It was difficult to use the equipment	2.15	1.11	0.72	0.79
CLEL4	I was overwhelmed by the features	2.30	1.23	0.67	0.81
CLEL5	I had problems to make observation	2.20	1.22	0.73	0.78
CLEL6	I had problems understanding the physics	2.40	1.28	0.62	0.83
	Average	**2.26**	**1.21**		
	Cronbach's α_C 95% Confidence interval		**0.85** [0.82, 0.87]		

Cognitive load X-ray experiment

All four items showed acceptable properties, and the scale showed good internal consistency with $\alpha_C = 0.84$, see Table 56.

Table 56. Item and scale properties of perceived cognitive load during the X-ray workshop rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

		Post (N = 507)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CLX1	It was difficult to use the equipment	2.05	1.10	0.73	0.78
CLX4	I was overwhelmed by the features	2.19	1.21	0.63	0.82
CLX5	I had problems to make observation	2.15	1.18	0.70	0.79
CLX6	I had problems understanding the physics	2.22	1.23	0.65	0.81
	Average	**2.15**	**1.18**		
	Cronbach's α_C 95% Confidence interval		**0.84** [0.82, 0.86]		

NASA Task Load Index (TLX)

In addition to the cognitive load scales described above, five items were adapted from the NASA Task Load Index (TLX) Hart (2006). Contrary to the usual agreement rating scale, students answered these five questions through a seven-level semantic differential. The lower endpoint was coded as 1, the neutral point as 4, and the higher endpoint was coded as 7. Due to the different rating options, this scale was analysed separately.

In the pilot study, this scale showed an acceptable internal consistency with $\alpha_C = 0.73$. However, in the main study the internal consistency was poor with $\alpha_C = 0.57$, see Table 57. The scale was originally designed to cover multiple dimensions of cognitive load, which

can explain a low internal consistency. Indeed, when comparing the mean values of the different items, it seems that students perceived the number (CL5), duration of the experiments (CL4) and their effort (CL5) as optimal with a neutral value around 4. However, the other two items covering the level of mental stress (CL1) and frustration (CL3) showed much lower values. Since cognitive load was already measured through the suitable scale "cognitive load in S'Cool LAB in general', the NASA task load scale was not used in further analyses.

Table 57. Item and scale properties of perceived task load rated on a seven-level semantic differential: Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale and the reduced item set including a 95% confidence interval.

		Post (N = 509)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CL1	How high was the level of mental stress you experienced in S'Cool LAB? (very low (1) to very high (7))	2.70	1.45	0.50	0.41
CL2	How hard did you have to work to perform the experiments? (very little(1) to very much (7))	3.90	1.34	0.34	0.51
CL3	How frustrated were you during your time in S'Cool LAB? (very little (1) to very much (7))	2.11	1.33	0.43	0.46
CL4	How do you rate the duration of the different experiments? (very short (1) to very long (7))	3.99	1.15	**0.17**	**0.60**
CL5	How do you rate the number of different experiments you were working with? (not enough (1) to too many (7))	3.39	1.06	**0.23**	0.57
	Cronbach's α_C 95% Confidence interval	**0.57** [0.51, 0.63]			

Summary of cognitive load in general and for the different workshops

Overall, students perceived their cognitive load in S'Cool LAB in general and related to the different experimental workshop to be very low. The workshop with the lowest cognitive load was the cloud chamber workshop (19% of the scale maximum), followed by the X-ray workshop (23% of the scale maximum) and the electron tube workshop (25% of the scale maximum). Interestingly, the item related to the germane cognitive load of the different workshops (*"I had problems understanding the physics principles of this experiment"*) was consistently rated highest across all three workshops. This suggests that the activities in S'Cool LAB were well-designed, as they provide students with enough support to handle the high-tech equipment and allow them instead to concentrate on the underlying physics principles. Furthermore, students perceived the number (three) and duration of the experiments (90 minutes each) as almost optimal, with a slight tendency towards wanting more experiments.

Although all four of the five subscales showed suitable properties, in the following, only the perceived cognitive load in general is used as a predictor in linear models of the educational outcomes. The other subscales are used to inform the continuous development of S'Cool LAB instead.

7.3.7 Cognitive activation

Students rated their agreement with five statements about their perceived cognitive activation in the post-visit questionnaire from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). Factor analysis confirmed the unidimensionality of the scale, all five items showed acceptable properties and the scale showed good internal consistency with $\alpha_C = 0.76$, see Table 58. The mean values of the scale were z-transformed and compared to a normal distribution. Here, the distribution of the z-values was significantly negatively skewed. The heavy tail of the distribution includes two extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution, these cases were not handled as outliers. Appendix 10.7.13 lists details about the distribution and extreme values.

Table 58. Item and scale properties of perceived cognitive activation rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

		Post (N = 509)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
CA1	I was focused when doing the experiments	5.06	0.85	0.54	0.71
CA2	I was actively involved in doing the experiments	5.28	0.76	0.58	0.70
CA3	I took a critical look at new ideas or material	4.66	1.05	0.55	0.71
CA4	I tried to distinguish between important and unimportant things when doing the experiments	4.49	1.05	0.50	0.73
CA5	During the experiments, I tried to connect what I was learning with my previous knowledge	5.09	0.98	0.49	0.73
	Cronbach's α_C	**0.76**			
	95% Confidence interval	[0.72, 0.79]			

7.3.8 Tutor support

Students rated their agreement with six statements about the perceived support by tutors in S'Cool LAB in the post-visit questionnaire from 'disagree completely" (coded as 1) to 'agree completely' (coded as 6). Factor analysis confirmed the unidimensionality of the scale, all six items showed acceptable properties and the scale showed good internal consistency with $\alpha_C = 0.90$, see Table 59. The mean values of the scale were z-transformed and compared to a normal distribution. Here, the distribution of the z-values is negatively skewed and also showed a significant positive kurtosis value. Indeed, the

tutors support scale suffered from a very strong ceiling effect with a scale mean of 5.37 ($\cong 87\%$ of the scale maximum). Consequently, the heavy tail of the distribution included five extreme cases with z-values below -3.29. However, due to the deviation from a normal distribution, these cases were not handled as outliers. Appendix 10.7.19 lists all details about the distribution and extreme values. In general, students highly appreciated the volunteering CERN scientists and perceived their support as very valuable.

Table 59. Item and scale properties of perceived tutor support rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item mean (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

		Post (N = 509)			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}
TS1	I had the chance to ask the tutors questions	5.33	0.81	0.70	0.88
TS2	I had the feeling tutors are fascinated by physics	5.50	0.78	0.66	0.89
TS3	The discussions with the tutors were important	5.16	0.94	0.69	0.88
TS4	I liked the working atmosphere in S'Cool LAB	5.39	0.83	0.75	0.87
TS5	The tutors responded to my questions	5.40	0.84	0.78	0.87
TS6	The tutors helped me with problems	5.43	0.82	0.76	0.87
	Cronbach's α_C	**0.90**			
	95% Confidence interval	[0.88, 0.91]			

7.3.9 Conceptual understanding

In total, students responded to seven multi-tier multiple-choice items in the pre- and the post-visit questionnaire. These items were adapted for the main study based on the results of the pilot study, see section 6.4. In particular, students did not answer any open text questions but chose their answers among different options. Also, the item texts were simplified to better adapt to the needs of students with limited English skills. Students answered up to three questions per multi-tier item, and their answers were graded and combined to an overall score per multi-tier item. In particular, some multi-tier items were graded as 50% correct if students answered at least partially correct. Appendix 10.7.10 includes the detailed grading rules, as well as the final item texts and an overview of students' answer behaviour in the pre- and post-visit questionnaire.

For this scale, the post-visit questionnaire was used for factor analysis because a more stable factor structure was expected, since students had learned more about the topic. The KMO value of 0.65 was comparatively low but still acceptable for factor analysis. Inspection of the scree plot after an exploratory principal axis factoring suggested three factors, see Figure 22, all with eigenvalues greater than 1.

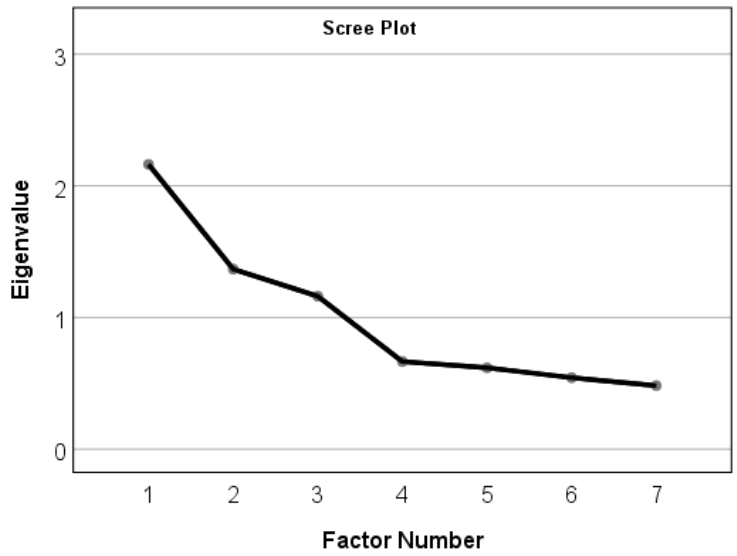


Figure 22. Scree plot of students' conceptual understanding as reported in the post-visit questionnaire.

Following a Varimax rotation with Kaiser Normalization (converging after 4 iterations), factor loadings ≥ 0.32 are displayed in Table 60 for all 7 items.

Table 60. Results of principal axis factoring with Varimax rotation & Kaiser Normalization. Factor loadings <0.32 are not shown.

		Factor		
Label	**Short description of the item**	**1**	**2**	**3**
PD	Proton mowing downwards	0.62		
PR	Proton mowing to the right	0.76		
PS	Proton at rest	0.58		
CM	Particle detector on a mountain		0.66	
CS	Particle detector in space		0.67	
XC	Dangers of X-ray cleaning			0.58
XW	Dangers of X-raying a watch			0.57
	Modelled variance:	**19%**	**13%**	**10%**

Together, the three extracted factors accounted for 42% of the variance in the variables and confirmed the expected factor structure:

- Items related to the electron tube workshop, Lorentz force (factor 1, accounts for 19% of the variance after Varimax rotation),
- Items related to the cloud chamber workshop, cosmic particles (factor 2, accounts for 13% of the variance) and
- Items related to the X-ray workshop, irradiation vs. contamination (factor 3, accounts for 10% of the variance).

Overall, the seven items showed acceptable properties, and the complete scale showed an acceptable internal consistency with $\alpha_C = 0.62$ for the pre-visit questionnaire and

$\alpha_C = 0.61$ for the post-visit questionnaire, see Table 61. Also, the sub-scales for the electron tube workshop and the cloud chamber workshop demonstrated acceptable internal consistency even with a very small number of items each. Only the subscale for the X-ray workshop showed a weak internal consistency with $\alpha_C = 0.44$ for the pre-visit questionnaire and $\alpha_C = 0.50$ for the post-visit questionnaire.

Table 61. Item and scale properties of students' conceptual understanding in total, and separately for the three workshops, rated on a six-level scale from 'disagree completely' (1) to 'agree completely' (6): Item easiness (M), standard deviation (SD), corrected item-scale correlation (r_{tj}), Cronbach's alpha when omitting item j (α_{tj}), Cronbach's alpha of the full scale including a 95% confidence interval.

Electron tube workshop		**Pre (n = 409)**				**Post (n = 453)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
PD	Proton mowing downwards	0.33	0.39	0.45	0.66	0.46	0.41	0.51	0.61
PR	Proton mowing to the right	0.32	0.47	0.55	0.53	0.34	0.47	0.55	0.55
PS	Proton at rest	0.46	0.50	0.52	0.58	0.54	0.50	0.48	0.65
	Cronbach's α_C [95% Conf. interval]	**0.69** [0.63, 0.74]				**0.70** [0.64, 0.74]			
Cloud chamber workshop		**Pre (n = 452)**				**Post (n = 509)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
CM	Particle detector on a mountain	**0.27**	0.45	0.50		0.54	0.25	0.43	
CS	Particle detector in space	0.40	0.49	0.50		0.45	0.25	0.43	
	Cronbach's α_C [95% Conf. interval]	**0.67** [0.60, 0.72]				**0.61** [0.53, 0.67]			
X-ray workshop		**Pre (n = 452)**				**Post (n = 509)**			
Label	**Short description of the item**	M	SD	r_{tj}	α_{tj}	M	SD	r_{tj}	α_{tj}
XC	Dangers of X-ray cleaning	0.52	0.50	0.28		0.66	0.23	0.34	
XW	Dangers of X-raying a watch	0.34	0.48	0.28		0.74	0.19	0.34	
	Cronbach's α_C [95% Conf. interval]	**0.44** [0.32, 0.53]				**0.50** [0.41, 0.58]			
Across all workshops, 7 items		**Pre (n = 409)**				**Post (n = 453)**			
	Cronbach's α_C [95% Conf. interval]	**0.62** [56, 0.67]				**0.61** [0.55, 0.66]			

Indeed, the items seemed to be quite difficult for the students in the sample, especially in the pre-visit questionnaire, with the lowest easiness score for item CM with only 27% correct answers. However, based on their learnings during the S'Cool LAB hands-on session, students answered more often correctly in the post-visit questionnaire. Only two items showed merely minor pre-post differences (items PR and item CS). Here, students' answer behaviour is analysed in more detail and interpreted in subsection 7.4.4.

Students' pre- and post-test scores were z-transformed and compared to a normal distribution. Here, the distribution of the z-values of the pre-test was positively skewed, which reflected the difficulty of the items for students. In comparison, the distribution of the z-values of the post-test only showed a small negative kurtosis value. Both the pre- and post-visit scores did not show any outliers. Appendix 10.7.10 lists details about the distribution.

In summary, the items and scales presented above were suitable to capture changes in conceptual understanding caused by the S'Cool LAB hands-on session. The items even separated into the expected three scales, and two of the subscales showed acceptable internal consistency. However, only the total test score will be considered as an educational outcome in the following. Furthermore, these items only captured a small aspect of the conceptual understanding of particle physics. Here, more research is needed to develop a concept test with a broader scope, see chapter 8.

7.3.10 School grades mathematics, physics and English skills

Students reported the grade in their latest grade report for the school subjects mathematics, physics and English in the grading system used in their school. In addition, teachers were asked to explain the grading system at their school while they were in S'Cool LAB with their students to help transform students' grades into a percentage scale. Here, interval scaled grade scales were assumed as a rough estimation. Grades that could not be interpreted in this way were coded as missing data. All three scales showed a significant deviation from a normal distribution, in particular, all were negatively skewed with the mathematics grade demonstrating the highest negative skewness, see appendix 10.7.10 for details about the distributions. In total, the heavy tails of the distributions included seven extreme cases with extreme values < -3.29. However, these cases were not treated as outliers due to the deviation from a normal distribution. Nevertheless, the mathematics grade was discarded as a predictor of cognitive effects in favour of the physics grade, which showed a more suitable distribution and fewer extreme values. Also, the physics grade showed a higher correlation with students' post-visit conceptual understanding ($r_{UN3,PH} = 0.33,\ p = 4 \cdot 10^{-13}$) compared to the mathematics grade ($r_{UN3,MA} = 0.26,\ p = 3 \cdot 10^{-8}$). Finally, students' mathematics and physics grades were highly correlated ($r_{MA,PH} = 0.69,\ p = 6 \cdot 10^{-72}$).

The English grade offered additional challenges: Some of the students did not pursue English courses at higher grades. Also, native speakers usually followed English literature courses instead. Thus, their English grade was not a good predictor of their English skills. Instead, it underestimated their English skills compared to non-native speaking students. Therefore, English grade was dichotomised in the following way: All native speakers were coded as 1, because it was assumed that English as the working language in S'Cool LAB did not offer any additional cognitive load for them. Furthermore, all students with an English grade $\geq 80\%$ were coded as 1 as well, because it was assumed that they were able to understand the English worksheets and instructions without problems. Students who did not report an English grade because they did not take English courses and all

students with English grades below 80% were coded as 0 (39% of the students). This left only two missing data points for two students who did not report any school grades at all. The high correlation between students' self-reported English grade and the dichotomised English skill suggested that both variables assessed similar constructs ($r = 0.67$**).

Table 62 provides an overview of students' self-reported school grades and the dichotomised English skills variable. In general, the negatively skewed distribution of students' school grades in this sample provided another indicator of a bias towards high-achieving students.

Table 62. Self-reported school grades for maths, physics, English and dichotomised English skills.

Range	Mathematics	Physics	English	English skill dichotomised	%
0% - 50%	11%	14%	8%	0 ($< 80\%$ non-native speakers + missing English course)	39%
51% - 60%	7%	11%	7%		
61% - 70%	18%	20%	18%		
71% - 80%	19%	15%	18%		
81% - 90%	32%	27%	39%	1 ($\geq 80\%$ + native speakers)	61%
91% - 100%	13%	13%	11%		
N	503	504	480		507
missing	6	5	29		2
M	75%	72%	77%		
SD	18%	18%	14%		
Min	0%	0%	0%		
Max	100%	100%	100%		

7.3.11 Previous experiences with particle physics

In S'Cool LAB and at CERN, students were confronted with particle physics, a topic that is not always part of physics curricula. Hence, any previous experience with this topic might have reduced students' perceived cognitive novelty in S'Cool LAB. Therefore, students were asked about their previous experience with particle physics ("*How many hours did you spend on particle physics so far?"*) in the pre-visit questionnaire. They were provided with seven answer options from zero hours to more than 20 hours. Here, students answered two separate items related to experience 'during physics classes' and 'outside physics classes'. Students' answer choices were transformed from a time interval to the central value of the respective interval, see Table 63. Also, a weighted average was calculated across all students. On average, students spent 5.5 hours on particle physics during their physics classes and 3.3 hours outside of their classes before visiting CERN and S'Cool LAB. 20% of the students had no or a maximum of 1-hour experience with particle physics beforehand. However, more than 50% of the students reported having spent more than five hours on particle physics, predominantly during their physics classes.

Since out-of-school learning contributed significantly to students' previous particle physics experience, the sum of both inside and outside particle physics experiences was used for future analyses. Here, the distribution of the z-transformed sum scores was positively skewed and thus deviated significantly from a normal distribution. The heavy tail of the distribution included two extreme cases with z-values > 3.29, which were not treated as outliers due to the skewed distribution, see appendix 10.7.14.

Table 63. Students' experience with particle physics during and outside of physics classes.

Time spent	Transformed to the centre of the interval	During class	Outside class	SUM
0 h	0 h	17%	26%	**9%**
0 – 1 h	0.5 h	16%	18%	**11%**
1 – 2 h	1.5 h	14%	19%	**16%**
3 – 5 h	4 h	18%	18%	**11%**
6 – 10 h	8 h	16%	10%	**20%**
10 – 20 h	15 h	15%	7%	**20%**
> 20 h	20 h	5%	2%	**13%**
	(Weighted) Mean	**5.5 h**	**3.3 h**	**8.8 h**
	Standard deviation	**6.0 h**	**4.7 h**	**8.9 h**

7.3.12 Previous experiences with hands-on experiments

In S'Cool LAB, students perform hands-on experiments in small groups. Although hands-on experimentation plays a central role in science education (Hofstein & Lunetta, 2004), the amount of experimentation in physics classrooms varies across countries (Börlin, 2012). Moreover, students with more hands-on experience in physics might feel more comfortable when doing experiments in S'Cool LAB and perceive a lower amount of capabilities-related novelty. Thus, students' answers to the question *"Experiments: How often do the following activities occur in your physics class?"* were analysed to capture this novelty influence factor.

Students were asked separately about the activities 'Individual work', 'Group work', 'Demonstration' and 'Simulation' and were provided with seven answer options from 'never' to 'every week'. Students' answer choices were transformed to a number of activities per year assuming 52 weeks per year and are summarised in Table 64. Also, a weighted average of the different activities per week was calculated across all students. On average, students experienced experimental activities 1.5 times per week with an equal amount of group work, demonstrations and simulations (approximately once every two weeks each) and a lower contribution from individual work (approximately once every five weeks).

These results are consistent with previous findings by Börlin (2012), who compared the frequency of different types of experiments in video recordings of 99 physics lessons from Germany, Finland and Switzerland. Similar to the results presented above, he reported that demonstrations by the teacher (58%) and group workshop (60%) were the most common types of experiments.

Table 64. Students' previous experience with experiments in their physics classes.

Answer option (code)	Transformed to # per year	Indiv. work	Group work	Demon-stration	Simu-lation
Never (1)	0	23%	6%	3%	4%
Very rarely (2)	1	23%	8%	9%	10%
Once every six months (3)	2	9%	8%	7%	8%
Once every two months (4)	6	10%	13%	10%	12%
Once a month (5)	12	12%	17%	24%	17%
Every two weeks (6)	26	16%	24%	27%	21%
Every week (7)	52	7%	25%	21%	28%
Weighted average number of experimental activities per week		**0.20**	**0.43**	**0.42**	**0.44**

Although demonstrations and simulations can certainly be useful to support learning, it is difficult to estimate their direct effect on students' perceived capabilities-related novelty when doing experiments on their own in S'Cool LAB. Therefore, only the sum of previous hands-on experiences (individual work and group work) was considered for the upcoming analyses. This sum was z-transformed and compared to a normal distribution. The distribution was positively skewed but did not include any extreme cases, see appendix 10.7.12.

To test by how much students from different countries varied in their hands-on experimentation experience, multi-level modelling was used to determine the variance partition coefficient for the level-2 (cluster) variables teacher, S'Cool LAB group, and school country. The results, which are summarised in Table 65, confirmed a very high amount of variance that can be attributed to the clustering of the data.

Table 65. Variance partition coefficients (VPC) for the level-2 (cluster) variables teacher, S'Cool LAB group and school country of the composed variable 'experience with hands-on experiments'.

	Teacher		Group		Country	
Model fit parameter AIC	1129		**1091**		1263	
Model fit parameter BIC	1137		**1100**		1272	
Variance components	σ^2	SE	σ^2	SE	σ^2	SE
Level-1 (student)	0.40	0.03	0.42	0.03	0.64	0.04
Level-2 (teacher/group/country)	0.74	0.16	0.58	0.17	0.38	0.16
VPC	**65%**		**58%**		**37%**	

In particular, 65% of the variance in students' experience with hands-on experimentation in their physics classes can be attributed to their physics teacher. The country of the students' school accounted for only 37% of the variance. However, the model fit parameters AIC and BIC are most favourable when using 'group' as the clustering variable. Here, students coming to S'Cool LAB together on the same date were considered one cluster, which accounted for 58% of the variance in students' hands-on experience. In some cases, one group coming to S'Cool LAB shared the same teacher. However, in other cases, students only shared the same school and were combined from different classes and teachers for their trip to CERN.

In summary, the country from which students came seemed to account for some of the differences in students' hands-on experimentation experience. However, the teacher of the students was a far more important factor. The cluster variable 'group' is a mix between the clustering variables 'teacher' and 'school' as is favoured by both model fit criteria.

7.3.13 Previous experience with out-of-school learning opportunities

If students had never been to an out-of-school learning lab like S'Cool LAB, it might take some time before they feel comfortable in this special learning setting because they may be confronted by a high amount of setting-familiarity-related novelty. If students, on the other hand, spend time in similar out-of-school learning places on a regular basis, they should be able to concentrate on the learning activities in S'Cool LAB more easily.

To assess students' prior experience with out-of-school learning opportunities, in particular, with science-related opportunities, they were asked to answer the following question in the pre-visit questionnaire together with the answer options provided in Table 66: *"Out-of-school learning places: You can not only learn things at school but also outside school – for example in museums and zoos. How often do you approx. visit the following places or use the following offers?"*

The seven answer options (ranging from 'never' to 'every week') were then transformed into a number per year (e.g., 12) while assuming 52 weeks per year. Afterwards, a weighted mean value was calculated for every student and out-of-school learning opportunity. On average, students taking part in the main study visited three museums per year, two zoos, aquariums or botanical gardens, two science centres or hands-on labs, and two universities or research labs. The by far biggest contribution to their out-of-school learning experience came from scientific TV shows (16 per year) and visits to science websites (19 per year), see Table 66.

Furthermore, 120 of the 509 students taking part in this study (24%) added another out-of-school learning opportunity by choosing the 'other' category. Here, 52 students mentioned books, magazines or a library. Also, 50 students mentioned sources in the internet including YouTube (42), MOOCs (3), social media (3), and TED talks (3). Additionally, 9 students mentioned their home, family or friends, 5 students mentioned extra-curricular courses and 4 students mentioned science camps. Indeed, web-based sources, in particular science YouTube channels, contributed significantly to students' science learning. This trend is also confirmed by the PISA 2006 data: *"It therefore seems that print and television media have the most influence over students in communicating information about science beyond the classroom. Additionally, in the majority of OECD countries a higher proportion of students reported that they regularly visited websites about science topics than borrowed or bought books about science topics"* OECD (2007).

Table 66. Students' experience with selected out-of-school learning opportunities.

Answer options (code)	Transformed to number per year	Museum	Zoo, aquarium or botanical garden	Science Centre or hands-on lab	University or research laboratory	Scientific TV shows	Science websites
Never (1)	0	14%	17%	25%	40%	14%	9%
Very rarely (2)	1	34%	46%	43%	36%	14%	16%
Once every six months (3)	2	29%	27%	22%	13%	11%	9%
Once every two months (4)	6	15%	6%	6%	5%	12%	11%
Once a month (5)	12	6%	3%	3%	3%	18%	16%
Every two weeks (6)	26	1%	1%	1%	2%	15%	18%
Every week (7)	52	1%	0%	0%	1%	17%	21%
Average number of the use of out-of-school learning opportunities per year		**3**	**2**	**2**	**2**	**16**	**19**

While web-based out-of-school science education was very popular among students, its influence on students' novelty experience in S'Cool LAB is difficult to estimate. Therefore, only the four 'real-world' categories of out-of-school science learning were considered and collapsed into one variable to obtain a suitable novelty influence factor for S'Cool LAB and a potential predictor of its educational outcomes. In particular, a weighted sum of students' scores for visits to the opportunities 'Museum', 'Zoo, aquarium or botanical garden', 'Science Centre or hands-on lab', and 'Universities or research laboratories' was formed, see Table 67.

Table 67. Number of visits to science-related out-of-school learning opportunities per year.

Number of science-related OSLEO visit per year	% students
0	5%
1 – 3	25%
4 – 6	32%
7 – 9	10%
10 – 12	9%
13 – 24	10%
> 24	10%
Average per year: 8	

Cors (2016) also included ‘Scientific TV shows’ and ‘Science websites’ in her collapsed technology-related out-of-school learning experience variable. Indeed, this variable was an important predictor of, for example, technology-related affective outcomes in her study. However, these two categories would dominate over the other categories in this sample. Also, it could be argued, that these two categories are rather indicators of science interest than a suitable measure of previous out-of-school learning experience relevant for educational outcomes at an out-of-school learning lab.

The distribution of the collapsed variable ‘experience with out-of-school science learning opportunities’ (EO) deviated significantly from a normal distribution. The distribution of z-scores was positively skewed and also showed a positive kurtosis value. The heavy tail of the distribution included nine cases with extreme values. However, due to the deviation from a normal distribution eight of the cases were not handled as outliers. Only case 410 had a z-score of 13.4 translating into 208 visits of out-of-school learning opportunities per year. This case was treated as an outlier and was adjusted towards the nearest extreme z-score of 5.1. Details about the extreme values and the distribution are included in appendix 10.7.13.

7.3.14 Summary of the items and scale analysis of the S’Cool LAB main study

In summary, the items that were included in the pre-visit and post-visit questionnaires of the S’Cool LAB main study showed suitable properties for further analysis. The scales’ dimensionalities were consistent with previous studies, and the selected subscales demonstrated good internal consistency. To measure cognitive effects in a very limited amount of testing time, seven concept-test-like items based on documented students’ conceptions with regard to the main topics of the three workshops were included in the pre- and post-visit questionnaires. The internal consistency of this scale was somewhat lower but still suitable to measure learning gains when keeping its limitations in mind. Indeed, this scale’s properties would benefit from additional items.

However, the distributions of the mean values of almost all scales deviated significantly from normal distributions. Here, Table 68 summarises the variables assessed in the pre-, inter- and post-visit questionnaires as well as their mean values, standard deviations and properties of the distributions.

Table 68. Variable plan for the three questionnaires of the S'Cool LAB main study including Cronbach's α_C, scale mean value – mostly on a 6-level scale (M), standard deviation (SD), mean value in percent of maximum possible score (M%), skewness value (S), kurtosis values (K) and number of extreme values $|z| > 3.29$ (#ev). Significant skewness and kurtosis values are indicated by an asterisk. The eight educational outcomes are marked as dependent variables (DV), and novelty influence factors (NIF) and novelty experience factors (NEF) are marked accordingly. Additional items for S'Cool LAB's formative assessment are not included.

t_1 pre	Variables (NIFs)	α_C	M	SD	M%	S	K	# ev
IE1	Dispositional physics interest	0.87	4.29	0.94	66%	-0.66*	0.72^{ns}	5
SC1	Dispositional physics self-beliefs	0.88	4.32	1.00	66%	-0.68*	0.39^{ns}	4
CT	Curiosity trait	0.89	5.25	0.69	85%	-1.41*	4.51*	4
CS1	Curiosity state particle physics	0.88	4.88	0.95	78%	-1.24*	2.21*	5
UN1	Conceptual understanding	0.62	0.38	0.26	38%	0.48*	-0.59^{ns}	0
EN	English skills (dichotomised)				61%			
PH	School grade Physics		0.72	0.18	72%	-0.86*	0.67^{ns}	3
EH	Experience hands-on experiments		32 y^{-1}	29 y^{-1}		0.92*	0.11^{ns}	0
EO	Experience out-of-school learning		53 y^{-1}	47 y^{-1}		1.33*	3.10*	9
EP	Experience with particle physics		8.8 h	8.9 h		1.08*	0.44^{ns}	2
SEX	Gender (Femaleness)				36%			
AGE	Age		17 y	0.9 y		-0.13^{ns}	-0.68^{ns}	0
t_2 inter	Variables	α_C	M	SD	M%	S	K	# ev
IE2	Dispositional physics interest	0.90	4.17	0.98	63%	-0.68*	0.26^{ns}	0
SC2	Dispositional physics self-beliefs	0.88	4.26	0.97	65%	-0.51*	-0.16^{ns}	0
CS2	Curiosity state particle physics	0.89	5.01	0.87	80%	-1.17*	2.08*	5
t_3 post	Variables (DVs & NEFs)	α_C	M	SD	M%	S	K	# ev
IE3	Situational physics interest (DV)	0.87	4.84	0.84	77%	-0.90*	0.85^{ns}	3
SC3	Situational physics self-beliefs (DV)	0.83	4.70	0.77	74%	-0.73*	1.48*	2
CS3	Curiosity state particle physics (DV)	0.95	4.87	1.00	77%	-1.10*	1.28*	3
CA	Perceived cognitive activation (DV)	0.76	4.91	0.67	78%	-0.63*	0.59^{ns}	2
UN3	Conceptual understanding (DV)	0.61	0.54	0.26	54%	-0.03^{ns}	-0.86*	0
ENJ	Perceived enjoyment (DV)	0.86	5.32	0.66	86%	-1.21*	2.14*	3
EAS	Perceived easiness (DV)	0.81	5.03	0.63	81%	-0.51*	0.34^{ns}	3
HEL	Perceived helpfulness (DV)	0.84	4.93	0.83	79%	-0.83*	0.59^{ns}	2
TIN	Tinkering support (NEF)	0.81	5.29	0.59	86%	-1.05*	1.34*	5
SO	Setting orientation (NEF)	0.70	4.62	0.75	72%	-0.28^{ns}	-0.21^{ns}	0
PRO	Lack of problems (NEF)	0.78	4.43	1.09	69%	-0.67*	-0.03^{ns}	0
CLE	Cognitive load experiments (NEF)	0.85	2.38	0.89	28%	0.80*	0.94*	3
TS	Perceived tutor support (NEF)	0.90	5.37	0.68	87%	-1.89*	6.06*	5

Many scales were negatively skewed and included several extreme values in their heavy tails. Also, some of the scales suffered from a ceiling effect, which can be explained by the special sample of highly motivated and curious students. Nevertheless, thanks to the relatively large sample size, the data was considered suitable for the statistical data analysis methods described in chapter 5 when keeping the limitations in mind.

7.4 Pre-post effects of the S'Cool LAB hands-on session

Dependent-samples t-tests were performed to compare the dispositional variables of the pre-visit questionnaire with the situational variables of the post-visit questionnaire. Furthermore, independent-samples t-tests were calculated to compare the scores of female and male participants.

As summarised in subsection 7.3.14, most distributions of the scales' mean values deviated significantly from normal distributions. However, as discussed in section 5.4, t-tests are quite robust against violations of a normal distribution if the sample size is big enough (Muijs, 2004). To compare the effects with similar studies, effects sizes (Cohen's d) were calculated if the t-tests confirmed significant differences.

Here, Table 69 summarises the impact of S'Cool LAB hands-on sessions on students' physics interest, self-beliefs, curiosity and their conceptual understanding.

Table 69. Differences between pre- and post-visit scores for the four educational outcomes interest, self-beliefs, curiosity and conceptual understanding in the main study ($N = 509$). The table includes the scale mean (M), standard deviation (SD), mean in percent of maximum possible score ($M\%$), and the results of the t-test including the correlation between pre- and post-test scores (r), t-statistic (t), p-value (p), effect size Cohen's d (d_C) and 95% confidence interval. Significance levels adjusted via Bonferroni correction ($m = 12$).[10]

	M	SD	M%	t-test results
Interest in physics				
Dispositional interest (pre)	4.29	0.94	**66%**	$r = 0.49, t(508) = -13.7, p = 10^{-37}$
Situational interest (post)	4.84	0.84	**77%**	$\boldsymbol{d_C = 0.6}$*****,** 95% CI $[0.5, 0.7]$
Physics self-beliefs				
Dispositional self-beliefs (pre)	4.32	1.00	**66%**	$r = 0.38, t(508) = -8.6\ p = 10^{-16}$
Situational self-beliefs (post)	4.70	0.77	**74%**	$\boldsymbol{d_C = 0.4}$*****,** 95% CI $[0.2, 0.5]$
Curiosity towards particle physics				
Curiosity state (pre)	4.88	0.95	78%	$r = 0.57, t(508) = 0.2, p = 0.82$
Curiosity state (post)	4.87	1.00	77%	
Conceptual understanding $(N = 380)$				
Score (pre)	0.39	0.26	**39%**	$r = 0.55, t(379) = -12.9, p = 10^{-31}$
Score (post)	0.56	0.26	**56%**	$\boldsymbol{d_C = 0.7}$*****,** 95% CI $[0.6, 0.8]$

Furthermore, Table 70 shows the comparison of students' physics interest, self-beliefs, curiosity and their conceptual understanding with respect to their gender.

[10] $m = 12$: * $p < 0.005$, ** $p < 0.001$ and *** $p < 0.0001$

Table 70. Gender differences for the four educational outcomes interest, self-beliefs, curiosity and conceptual understanding in the main study ($N = 509$). The table includes the scale mean (M), standard deviation (SD), mean in percent of maximum possible score ($M\%$), and the results of the t-test including the t-statistic (t), p-value (p), effect size Cohen's d (d_C) and 95% confidence interval. Significance levels adjusted via Bonferroni correction ($m = 12$).

	Girls ($n = 182$)			**Boys** ($n = 327$)			**t-test results**
	M	**SD**	**M%**	**M**	**SD**	**M%**	(comparison boys – girls)
Interest in physics							
Dispositional interest (pre)	4.11	0.98	**62%**	4.39	0.90	**68%**	$t(347) = 3.2, p = 0.002$ $\boldsymbol{d_C = 0.3^{*}}$; 95% CI $[0.1, 0.5]$
Situational interest (post)	4.87	0.84	77%	4.82	0.83	76%	$t(507) = -0.6, p = 0.53$
Physics self-beliefs							
Dispositional self-beliefs (pre)	4.08	1.01	**62%**	4.46	0.97	**69%**	$t(507) = 4.2, p = 4 \cdot 10^{-5}$ $\boldsymbol{d_C = 0.4^{***}}$; 95% CI $[0.2, 0.6]$
Situational self-beliefs (post)	4.62	0.85	72%	4.75	0.73	75%	$t(507) = 1.7, p = 0.10$
Curiosity towards particle physics							
Curiosity state (pre)	4.80	0.94	76%	4.92	0.95	78%	$t(507) = 1.4, p = 0.16$
Curiosity state (post)	4.95	1.01	79%	4.83	0.99	77%	$t(507) = -1.4, p = 0.18$
Conceptual understanding	$n_{pre} = 140$ $n_{post} = 149$			$n_{pre} = 269$ $n_{post} = 304$			
Score (pre)	0.36	0.25	36%	0.39	0.26	39%	$t(407) = 1.3, p = 0.18$
Score (post)	0.50	0.24	50%	0.56	0.26	56%	$t(451) = 2.4, p = 0.02$

7.4.1 Interest in physics

The dependent-samples t-test results confirmed that the mean value of students' triggered situational interest was significantly higher than the mean value of students' dispositional interest in physics. This difference translates into a medium-sized effect with Cohen's $d_C = 0.6$, see Table 69. Figure 23 shows the pre-post-values of the respective variables in per cent for the whole group, and separately for girls and boys. Indeed, the mean value of the dispositional interest of girls was significantly lower than the dispositional interest of boys (small effect, $d_C = 0.3$), but there was no significant difference in their triggered situational interest, see Table 70. Therefore, with respect to their physics interest, participation in the S'Cool LAB hands-on session had a stronger effect on girls $d_C = 0.8$ (95% CI $[0.6, 1.0], r = 0.48, t(181) = -10.9, p = 10^{-21}$) than on boys $d_C = 0.5$ 95% CI $[0.3, 0.6], r = 0.51, t(326) = -9.0, p = 10^{-17}$).

7.4.2 Physics self-beliefs

When comparing the scores for triggered situational self-beliefs and students' dispositional physics self-beliefs, the results of a dependent-samples t-test confirmed a significant

difference translating into a small effect with Cohen's $d_C = 0.4$, see Table 69. Figure 24 shows the pre-post-values of the respective variables in per cent for the whole group, and separately for girls and boys. Indeed, there was a small difference between girls' and boys' dispositional self-beliefs ($d_C = 0.4$) but no significant difference in their triggered situational self-beliefs,. Therefore, with respect to their physics self-beliefs, participation in the S'Cool LAB hands-on session had a stronger effect on girls $d_C = 0.5$ ($95\%\ CI\ [0.3, 0.7], r = 0.36, t(181) = -6.9, p = 8 \cdot 10^{-11}$) than on boys $d_C = 0.3$ $95\%\ CI\ [0.1, 0.4], r = 0.39, t(326) = -5.5, p = 10^{-8}$).

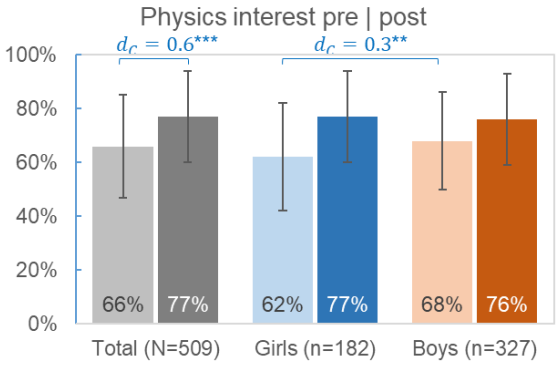


Figure 23. Significant difference between dispositional interest & situational interest and significant difference between girls & boys in their dispositional interest. Error bars indicate ±1 standard deviation.

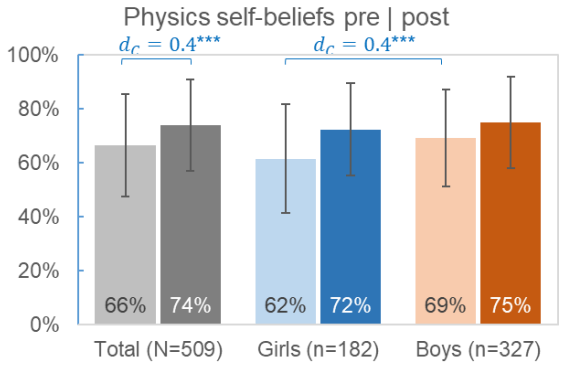


Figure 24. Significant difference between dispositional self-beliefs & situational self-beliefs and significant difference between girls & boys in their dispositional interest. Error bars indicate ±1 standard deviation.

7.4.3 Curiosity towards particle physics

Students were asked about their curiosity towards particle physics in the pre-visit questionnaire (three items) and the post-visit questionnaire (nine items). Here, dependent-samples t-tests did not identify significant differences between the two measurements, see Table 69. Also, there was no significant difference between female and male participants, see Table 70. Indeed, the level of students' curiosity towards particle physics was already very high in the pre-visit questionnaire (78% of the scale maximum), which left little room for improvement. Figure 25 shows the pre-post-values of the respective variables in per cent for the whole group, and separately for girls and boys.

7.4.4 Conceptual learning

Students answered seven multi-tier multiple-choice items before and after taking part in the S'Cool LAB hands-on session. The items were selected and equipped with distractors based on the results of the pilot study, see subsection 6.4.4. In the pre-visit questionnaire, these test items were challenging for students with an average test score of 39%. However, students improved their test-score significantly in the post-visit questionnaire to 56%. This

pre-post difference translates into a medium-sized effect with $d_C = 0.7$, see Table 69. When comparing test-scores of girls and boys, no significant differences were identified, see Table 70. Figure 26 shows the pre-post-values of the respective variables in per cent for the whole group, and for separately girls and boys.

The use of multi-tier test items does not just allow to calculate an overall test-score, but also to analyse students' answer behaviour and reasoning patterns as well as changes in their answers from the pre- to the post-visit questionnaire. Therefore, interesting aspects of these seven items are reported separately in the following. Moreover, appendix 10.7.10 includes the detailed grading rules for the multi-tier items, the final item texts and the details of students' answers in the pre- and post-visit questionnaire.

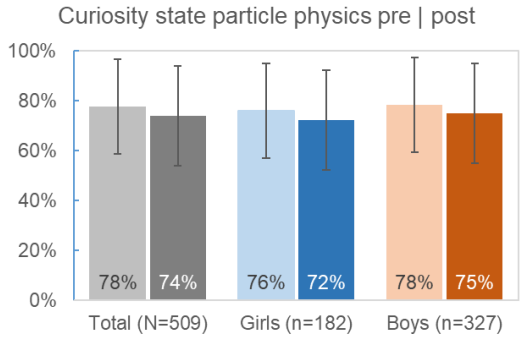


Figure 25. Curiosity development: no significant effects. Error bars indicate ±1 standard deviation.

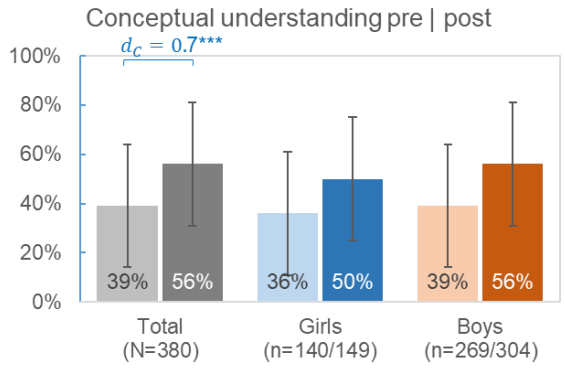


Figure 26. Significant difference between the pre- and post-test scores of students' conceptual understanding. Error bars indicate ±1 standard deviation.

7.4.4.1 Electron tube workshop: the Lorentz force and protons in magnetic fields

Understanding the behaviour of electrically charged particles in magnetic fields is a key concept of the electron tube workshop in S'Cool LAB. Hence, students were asked to answer three 4-tier items during the pre- and post-visit questionnaire to assess their understanding of this concept. In particular, they were confronted with examples of a proton moving perpendicular to a uniform magnetic field (item PD), of a proton moving along the field lines of a uniform magnetic field (item PR), and of a proton at rest in a uniform magnetic field (item PS). For each situation, they had to decide whether the magnets exert a force on the proton (tier-1), and if yes, decide in which direction this force is directed (tier-2). Furthermore, they had to choose one of six answer options to explain their reasoning (tier-3) and indicated their confidence with their answers (tier-4).

Item PD: A proton is moving downwards (perpendicular) through a magnetic field

Tier-1 (force yes/no): In the **pre-visit questionnaire**, almost all students (91%) answered correctly that the magnets exert a force on a proton moving perpendicular through a

magnetic field. The fraction of correct answers increased only slightly to 92% for the **post-visit questionnaire**.

Tier-2 (direction): However, in the **pre-visit questionnaire**, only 35% of these students who expected a force also correctly identified the direction of the force. Here, 6% of the students who expected a force correctly identified a force perpendicular to the magnetic field and the movement of the proton but chose the exact opposite direction[11]. Furthermore, 31% of the students expected a force along the magnetic field lines either to the right (towards the south pole of the magnets, 19%) or to the left (towards the north pole of the magnets, 12%).

In the **post-visit questionnaire**, students' answer behaviour changed in a favourable way. Now, 39% of the students who expected the magnets to exert a force on the proton also correctly identified the direction of the force. Further 17% chose the exact opposite direction. However, 18% of the students expected a force along the magnetic field lines (12% towards the south pole, 6% towards the north pole of the magnets). All other directions were chosen significantly less (<10% of the students per option).

Tier-3 (explanation): In the **pre-visit questionnaire**, 30% of the students argued with the right-hand-rule. However, 13% argued that '*the south pole attracts the proton'*, 11% said '*the north pole attracts the proton'* and 18% argued that '*the force acts in the direction of the magnetic field*'. Also, 17% claimed that none of the provided explanations applies.

In the **post-visit questionnaire**, most students (51%) argued with the right-hand-rule. This is no surprise since the right-hand-rule was majorly featured and used during the electron tube workshop in S'Cool LAB. Also, 16% of the students answered that the poles of the magnet attracted the proton (9% attracted by south pole, 7% attracted by north pole) and another 15% argued with a force in the direction of the magnetic field. Besides, 11% claimed that none of the provided explanations applies.

Tier-4 (confidence): Overall, students increased their confidence in their answers from 44% to 53% in the post-visit questionnaire.

Figure 27 displays a **Sankey diagram with the changes in students' answer behaviour from the pre- to the post-visit questionnaire**. Many students continued to expect a force in the post-visit questionnaire, and most of them chose the right direction or the exact opposite direction. However, several students who changed from 'no' to 'yes' at tier-1

[11] To reduce the difficulty of this item, both perpendicular directions were graded as 'correct', see Appendix 10.7.10 for the detailed grading rules.

chose a wrong direction of the force at tier-2. Also, 6% of the students changed their answer in an unfavourable way from 'yes' to 'no' at tier-1.

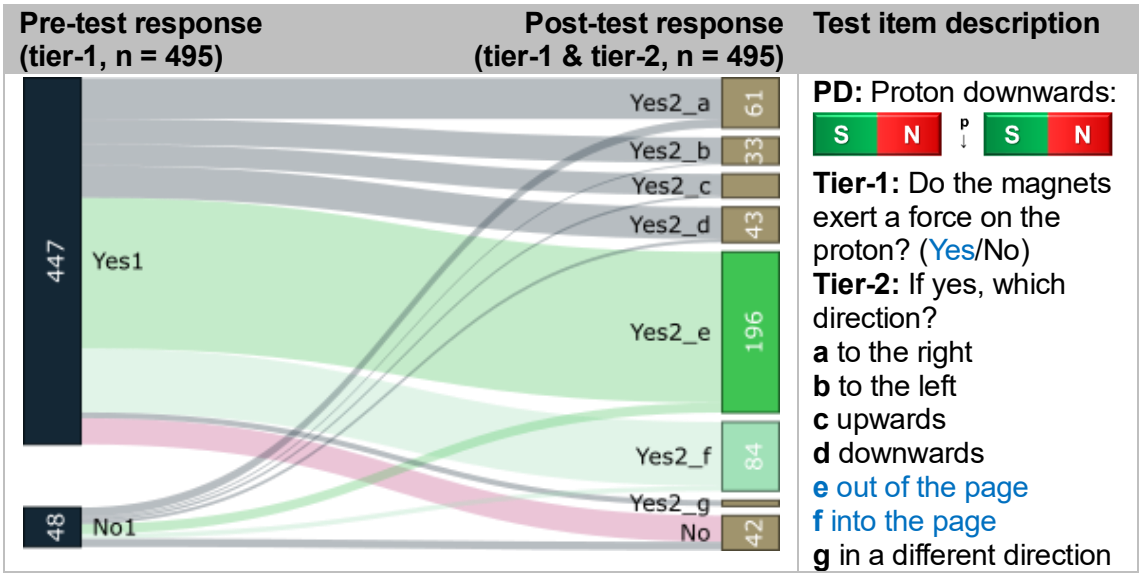


Figure 27. Sankey diagram: A proton is moving downwards (perpendicular) through a uniform magnetic field (item PD). The diagram shows students' answers to the tier-1 level of the item (force yes/no) before taking part in the S'Cool LAB hands-on session (pre-test, left side), and their answers to the tier-1 (force yes/no) and tier-2 (direction) level of the item afterwards (post-test, right side).

Additional information can be derived from the **combinations of students' answers at tier-2 (direction of the force) and tier-3 (explanation).** Here, the percentage of students choosing each combination is listed in Table 137 in appendix 10.7.10. In summary, most students answered consistently across different item levels and their reasoning can be assigned either to the correct solution or to one of the documented students' conceptions. For example, in the pre-visit questionnaire, 10% of the students chose the direction to the right (towards the south pole) together with the explanation '*the south pole attracts the proton*' and 9% chose the direction to the left (towards the north pole) together with the explanation '*the north pole attracts the proton*'. Both answer combinations can be interpreted in the framework of the 'charged poles' conception or the 'attractive/repulsive poles' conception described by Maloney et al. (2001). However, these answer combinations were chosen less frequently in the post-visit questionnaire with only 6% (south pole) and 3% (north pole), respectively.

Furthermore, in the post-visit questionnaire, most students argued with the right-hand-rule. However, several students still had problems to apply the rule in the right way and chose the wrong direction (tier-2) together with the correct explanation (tier-3). Indeed, students often chose the direction opposite to the correct direction when applying the right-hand-rule, which is consistent with results documented by Li and Singh (2017). For this item, 5% argued with the right-hand-rule in the pre-visit questionnaire, but chose the opposite direction. In the post-visit questionnaire, this answer combination was chosen by even 14%

of the students. Here, it seems that students correctly learned to argue with the right-hand-rule as an effect of the workshop in S'Cool LAB but did not fully understand how to apply the right-hand-rule correctly. However, 18% in the pre-visit questionnaire and 29% in the post-visit questionnaire chose the correct answer combination including the correct direction.

Last, 13% of the students in the pre-visit questionnaire (and 8% in the post-visit questionnaire) chose the correct direction together with the explanation option 'none of the explanation above applies'. One possible interpretation could be that these students learned about the Lorentz force in school in a different way, i.e., without the right-hand-rule or with a differently named rule.

Item PR: A proton is moving to the right (along the field) through a magnetic field

Tier-1 (force yes/no): In the **pre-visit questionnaire**, 47% of the students correctly answered that the magnets do not exert a force on a proton moving to the right in the direction of a uniform magnetic field. This percentage increased slightly to 51% for the **post-visit questionnaire**.

Tier-2 (direction): Among the 53% of the students who incorrectly expected a force in the **pre-visit questionnaire**, 53% expected a force to the right, and 25% a force to the left. In the **post-visit questionnaire**, only 34% of the students who expected a force indicated that this force would point to the right, and 17% to the left. Instead, for the rest of the students who expected a force, the choice of the force's directions was more or less equally distributed. In the post-visit questionnaire, 16% chose upwards, 10% downwards, 9% out of the page, 10% into the page, and 4% in a different direction.

Tier-3 (explanation): In the **pre-visit questionnaire**, 30% answered that the magnetic poles attract the proton (south pole 18%, north pole 12%), while 15% said that the '*force acts in the direction of the magnetic field*'. Also, 17% of the students answered that the '*force acts in the direction of the movement of the proton*'. These explanations were consistent with students' preferred choices of the force's direction. Furthermore, 13% of the students argued with the right-hand-rule. Finally, 25% of the students said that none of the explanations would apply.

In the **post-visit questionnaire**, more students (27%) argued with the right-hand-rule while the other explanations were chosen with similar frequencies as in the pre-visit questionnaire. However, fewer students argued with attracting magnetic poles (11% south pole, 8% north pole).

Tier-4 (confidence): Overall, students increased their confidence in their answers from 40% to 49% in the post-visit questionnaire.

Figure 28 displays a **Sankey diagram with the changes in students' answer behaviour from the pre- to the post-visit questionnaire**.

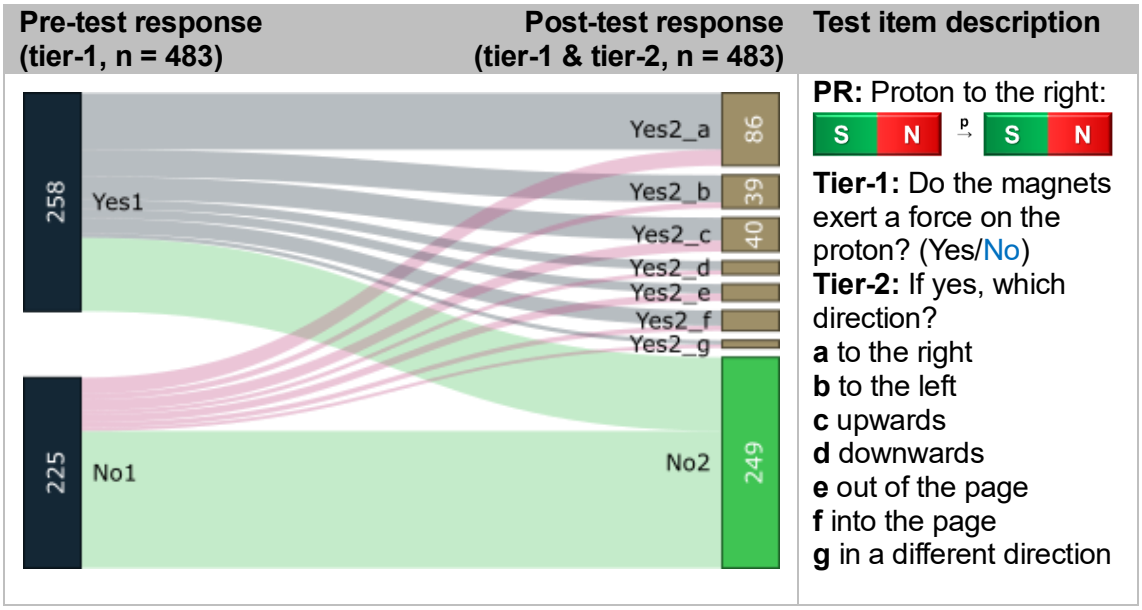


Figure 28. Sankey diagram: A proton is moving to the right along the magnetic field lines (item PR). The diagram shows students' answers to the tier-1 level of the item (force yes/no) before taking part in the S'Cool LAB hands-on session (pre-test, left side), and their answers to the tier-1 (force yes/no) and tier-2 (direction) level of the item afterwards (post-test, right side).

Many students continued to expect no force in the post-visit questionnaire. Interestingly, of the 225 students who answered correctly (Force: No) in the pre-visit questionnaire, 63 students changed their opinion to the wrong answer (Force: Yes) in the post-visit questionnaire. These students' choices of the force's direction were almost equally distributed across the six options. This trend can be considered an adverse effect of the electron tube workshop. However, of the 259 students who expected a force in the pre-test, 87 changed their opinion to the correct answer (no force) in the post-test. Therefore, the workshop has a net positive effect on students' answers at this item level.

Additional information can be derived from the **combinations of students' answers at tier-2 (direction of the force) and tier-3 (explanation)**. Here, the percentage of students for all answer combinations is listed in Table 139 in appendix 10.7.10. In summary, most students answered consistently across different item levels. In particular, 26% of the students in the pre-visit questionnaire (and still 16% post) chose the direction 'to the right' (towards the south pole) together with the explanation '*the south pole attracts the proton*'. Similarly, 16% in the pre-visit questionnaire (and 10% post) chose the direction 'to the left' (towards the north pole) together with the explanation '*the north pole attracts the proton*'.

Again, these reasoning patterns can be interpreted in the framework of the 'charged poles' conception or the 'attractive/repulsive poles' conception described by Maloney et al. (2001). Also, 15% pre (and 7% post) chose the direction 'to the right' together with the explanation that '*the force acts in the direction of the magnetic field*', which could be an indicator of students confusing electric and magnetic fields and their associated forces (Maloney et al., 2001) or students thinking that the proton only moves through the magnetic fields, because a force acts on it (Li & Singh, 2017).

The frequencies of these wrong answer patterns decreased from the pre- to the post-visit questionnaire. Instead, the frequency of combinations of the explanation 'right-hand-rule' and different directions increased (from 12% to 31% in total in the post-visit questionnaire). For example, 9% of the students in the post-visit questionnaire chose the wrong direction 'upwards' together with the right-hand-rule explanation. Indeed, several students seemed to have problems to apply the rule in the right way after the short intervention. This comes as no surprise as the electron tube workshop in S'Cool LAB introduced students to the right-hand-rule without explicitly discussing special cases, such as particles moving along the magnetic field lines.

Item PS: A proton is at rest in a uniform magnetic field

Tier-1 (force yes/no): In the **pre-visit questionnaire**, 54% students correctly answered that the magnets do not exert a force on a proton at rest in a uniform magnetic field. This increased to 62% for the **post-visit questionnaire**.

Tier-2 (direction): Among the 46% of the students who incorrectly expected a force in the **pre-visit questionnaire**, 46% expected a force to the right, and 24% a force to the left. In the **post-visit questionnaire**, only 38% of the students who expected a force indicated that this force would point to the right, and 22% to the left. Instead, for the rest of the students who expected a force the choice of the force's directions was more or less equally distributed. In the post-visit questionnaire, 7% chose upwards, 9% downwards, 12% out of the page, 4% into the page, and 8% in a different direction.

Tier-3 (explanation): In the **pre-visit questionnaire**, 25% answered that the magnetic poles attract the proton (14% south pole, 11% north pole), while 14% said that the '*force acts in the direction of the magnetic field*'. This is consistent with students' preferred choices of the force's direction at tier-2. Only 8% of the students argued with the right-hand-rule. Finally, 48% of the students explained that none of the explanations applies to a proton at rest.

In the **post-visit questionnaire**, more students (18%) argued with the right-hand-rule. Also, 44% of the students explained that for a proton at rest none of the explanations would apply. The other explanations were chosen similarly often by the students. However, fewer students argued with attracting poles (11% south pole attracts, 9% north pole attracts).

Tier-4 (confidence): Overall, students increased their confidence in their answers slightly from 40% to 46% in the post-visit questionnaire.

Figure 29 displays a **Sankey diagram with the changes in students' answer behaviour from the pre- to the post-visit questionnaire**. Many students continued to expect no force in the post-visit questionnaire. 99 students (21%) changed their answer from wrong (Force: Yes) in the pre-visit questionnaire to correct (Force: No) in post-test. However, of the 255 who answered correctly (Force: No) in the pre-test, 56 changed their opinion to the wrong answer (Force: Yes) in the post-visit questionnaire together with different choices of directions. This can be considered as a negative effect of the electron tube workshop. However, overall, the workshop had a net positive effect on students' answers to this item level.

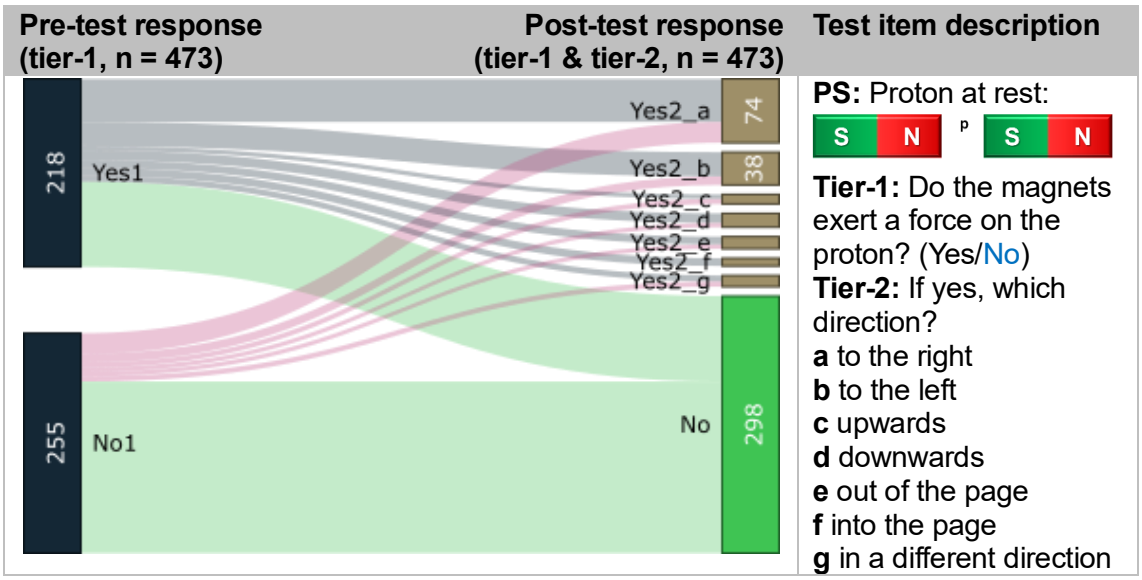


Figure 29. Sankey diagram: A proton is at rest in a uniform magnetic field (item PS). The diagram shows students' answers to the tier-1 level of the item (force: yes/no) before taking part in the S'Cool LAB hands-on session (pre-test, left side), and their answers to the tier-1 (force: yes/no) and tier-2 (direction) level of the item afterwards (post-test, right side).

Additional information can be derived from the **combinations of students' answers at tier-2 (direction of the force) and tier-3 (explanation)**. Here, the percentage of students for all answer combination is listed in Table 141 in appendix 10.7.10. Equally to the previous items, most students answered consistently across different item levels. In particular, 25% of the students in the pre-visit questionnaire (and still 2% post) chose the direction 'to the right' (towards the south pole) together with the explanation '*the south pole*

attracts the proton'. Similarly, 19% in the pre-visit questionnaire (and 14% post) chose the direction 'to the left' (towards the north pole) together with the explanation '*the north pole attracts the proton*'. As for the previous two items, these reasoning patterns can be interpreted in the framework of the 'charged poles' conception or the 'attractive/repulsive poles' conception described by Maloney et al. (2001). Also, 17% of the students in the pre-visit questionnaire (and 7% post) chose the direction 'to the right' together with the explanation that '*the force acts in the direction of the magnetic field*', which could be an indicator of students confusing electric and magnetic fields and their associated forces (Maloney et al., 2001).

As before, the frequencies of the wrong answer patterns decreased from the pre- to the post-visit questionnaire. Instead, the frequency of combinations with the explanation 'right-hand-rule' and different directions increased (from 9% to 17% in total in the post-visit questionnaire). However, 4% of the students in the post-visit questionnaire chose the wrong direction 'out of the page' together with the right-hand-rule explanation. This can be interpreted again as a side-effect of the electron tube workshop in S'Cool LAB, which introduced students to the Lorentz force and the right-hand-rule without explicitly discussing cases with particles at rest.

Furthermore, 8% of the students in the pre-visit questionnaire (4% post) chose either a different direction than the suggested six direction or 'none of the explanations applies'. Here, one possible explanation could be that students not only considered the force that can be exerted by the magnets, but also the gravitational force. In this case, they might have thought that the magnets attract the proton to the right while gravity acts downwards, which would result in a force in between those two directions. Here, further data from the students would be necessary to explain their reasoning.

7.4.4.2 Cloud chamber workshop: cosmic particles and particle detectors

To assess students understanding of cosmic particles, which is a key concept of the cloud chamber workshop in S'Cool LAB, students answered two 3-tier items during the pre- and post-visit questionnaire. In particular, they were asked to compare two particle detectors in two different scenarios: a particle detector on top of a mountain versus one in the valley (item CM) and a particle detector on a satellite versus one on the surface of the Earth (item CS). For both situations, students had to decide which detector would record a higher count rate and explain their reasoning. In addition, students indicated their confidence with their answers (tier-3).

Item CM: Particle detector on a mountain at 5000 m vs. in a valley at 0 m altitude

Tier-1 (comparison of detection rate): In the **pre-visit questionnaire**, 37% of the students answered correctly that there would be a higher number of particles detected on top of the mountains. This increased to 54% for the **post-visit questionnaire**.

Tier-2 (explanation): In the **pre-visit questionnaire**, 19% identified space as the source of the detected particles and 15% argued that the atmosphere has a shielding effect. Both answers were graded as correct. However, 40% of the students incorrectly answered that '*fewer air molecules on top of the mountain*' would cause the different detection rates of the particle detectors.

In the **post-visit questionnaire**, students' answer behaviour changed in a favourable way. Now, 37% of the students correctly argued with space as the source of the detected particles and 17% argued that the atmosphere has a shielding effect. However, even after the cloud chamber workshop, 20% of the students answered that '*fewer air molecules on top of the mountain*' would cause the different detection rates of the particle detectors. The other distractors were chosen by 5-10% of the students, respectively.

Tier-3 (confidence): Overall, students increased their confidence in their answers from 44% to 54% in the post-visit questionnaire.

Figure 30 displays a **Sankey diagram with the changes in students' answer behaviour from the pre- to the post-visit questionnaire**. 27% of the 509 students improved their answer behaviour at tier-1 from a wrong to the correct answer option. However, a significant number of students (26%) stayed with their wrong answer suggesting rather stable conceptions. Only a small fraction of students (11%) changed their answer in an unfavourable way from the correct to the wrong option.

Additional information can be derived from the **combinations of students' answers at tier-1 (comparison of detection rate) and tier-2 (explanation)**. Here, the percentage of students for all answer combination is listed in Table 143 in appendix 10.7.10. In summary, most students answered consistently across different item levels and their reasoning can be assigned to the correct solution or to one of students' documented conceptions. For example, 37% of the students in the pre-visit questionnaire (and still 16% in the post-visit questionnaire) expected a lower count-rate on top of the mountain together with the explanation that there were '*fewer air molecules*'. Similarly, 7% of the students in the pre-visit questionnaire (8% post) answered that the particle detection rate would be the same because the properties of the atmosphere were the same for both detectors. Both answer combinations are consistent with the findings obtained within the pilot study (subsection

6.4.2) suggesting that students' alternative interpretation of 'particles' in this context can cause severe challenges. However, these answer combinations were surprisingly less frequent in the post-visit questionnaire. Last, 5% of the students in the pre-visit questionnaire (and 5% post) expected a lower number of detected particles because 'the Earth is the source of the detected particles'.

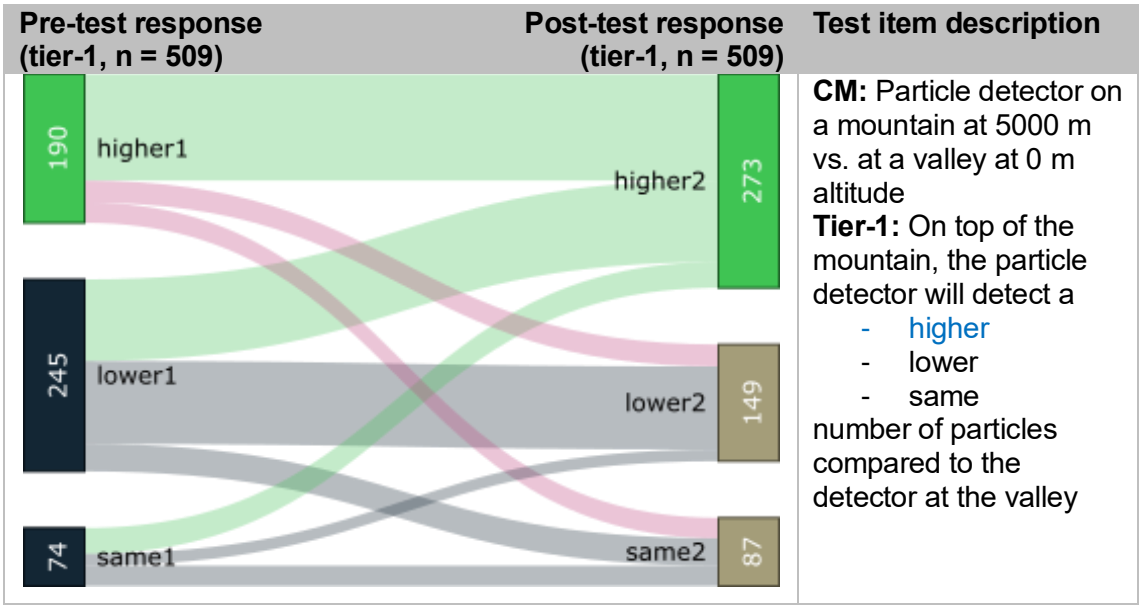


Figure 30. Sankey diagram: Particle detector on a mountain at 5000 m vs. at a valley at 0 m altitude (item CM). The diagram shows students' answers to the tier-1 level of the item before (pre-test, left side) and after (post-test, right side) taking part in the S'Cool LAB hands-on session.

Item CS: Particle detector on a satellite at 400 km vs. on Earth at 0 km altitude

Tier-1 (comparison of detection rate): In the **pre-visit questionnaire**, 47% of the students correctly answered that there would be a higher number of particles detected on the satellite in space. This increased to 64% for the **post-visit questionnaire**.

Tier-2 (explanation): In the **pre-visit questionnaire**, 24% of the students identified space as the source of the detected particles and 23% argued that the atmosphere has a shielding effect. Both these answers were graded as correct. However, 28% of the students answered that '*fewer air molecules in space*' would cause the different detection rates of the particle detectors. Further 16% of the students argued that the '*Earth is the main source of the detected particles*'. The other two distractors were chosen only by 3% and 5% of the students, respectively.

In the **post-visit questionnaire**, students' answer behaviour changed in a favourable way. Now, 38% of the students correctly argued with space as the source of the detected particles and 26% argued that the atmosphere has a shielding effect. After taking part in the cloud chamber workshop, 14% of the students still answered that '*fewer air molecules in space*' would cause the different detection rates of the particle detectors. Further 15%

of the students argued that the '*Earth is the main source of the detected particles*'. The other two distractors were chosen only by 3% and 4% of the students, respectively.

Tier-3 (confidence): Overall, students increased their confidence in their answers from 44% to 54% in the post-visit questionnaire.

Figure 31 displays a **Sankey diagram with the changes in students' answer behaviour from the pre- to the post-visit questionnaire**. As for item CM, many students (26%) improved their answer behaviour in a favourable way at tier-1 after taking part in the S'Cool LAB hands-on session. However, a significant number of students (23%) stayed with their wrong answer suggesting rather stable conceptions. Only 8% of the students changed their answer in an unfavourable way from the correct to the wrong option at tier-1.

Additional information can be derived from the **combinations of students' answers at tier-1 (comparison of detection rate) and tier-2 (explanation)**. The percentage of students for all answer combination is listed in Table 145 in appendix 10.7.10. In summary, most students answer consistently across different item levels and their reasoning can be assigned to the correct solution or to one of students' documented conceptions. For example, 27% of the students in the pre-visit questionnaire (and still 12% in the post-visit questionnaire) expected a lower count-rate on top of the mountain together with the explanation that there were '*fewer air molecules*'. Again, this answer combination is consistent with findings from the pilot study. Also, 13% of the students in the pre-visit questionnaire (8% post) answered that the particle detection rate would be lower because the 'Earth is the main source of the detected particles'.

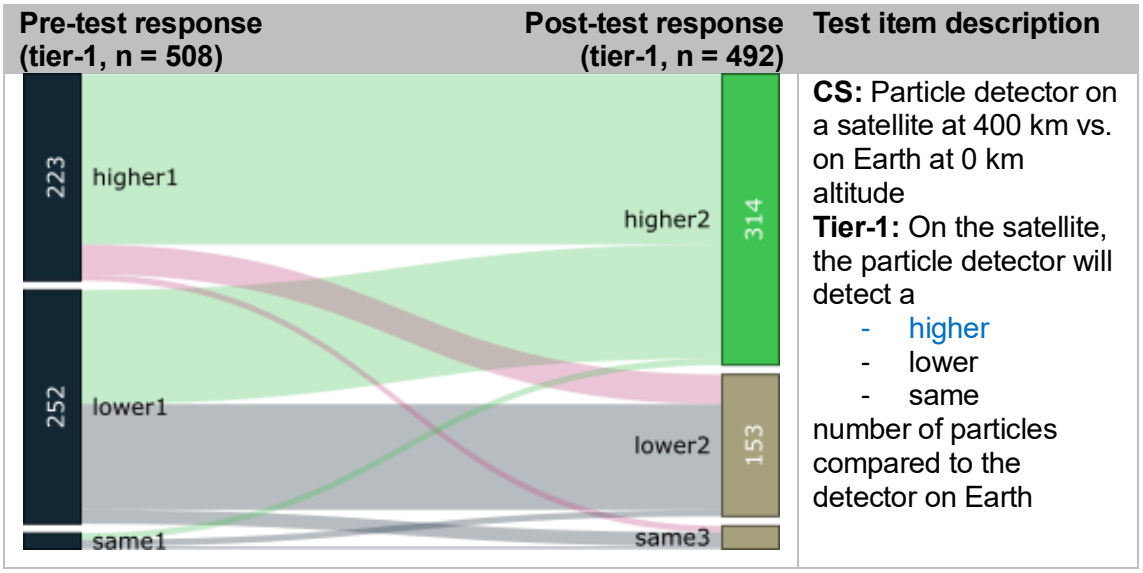


Figure 31. Sankey diagram: Particle detector on a satellite at 400 km vs. on Earth at 0 m altitude (item CS). The diagram shows students' answers to the tier-1 level of the item before (pre-test, left side) and after (post-test, right side) taking part in the S'Cool LAB hands-on session.

7.4.4.3 X-ray workshop: irradiation vs. contamination and X-radiation

To assess students' understanding of the interactions of X-radiation with matter, which is a key concept of the X-ray workshop in S'Cool LAB, students answered two 2-tier items during the pre- and post-visit questionnaire. In particular, they were asked to evaluate the dangers associated with cleaning a room with an X-ray machine (item XC) and the danger of wearing a watch that has been X-rayed for 1 hour (item XW). For both situations, students had to decide, whether or not the situation was dangerous and choose an explanation for their reasoning (tier-1). These two steps were combined in one level based on the item version used by Molz (2016): students answered two multiple-choice items with four or five answer options, which included a 'Yes' or 'No' together with an explanation. In addition, students rated their confidence with their answer (item-2).

Item XC: Dangers of cleaning a room with an X-ray machine

Tier-1: In the **pre-visit questionnaire**, 52% of the students answered correctly that there is no danger when cleaning a room with an X-ray machine if the X-ray units are switched off. In the **post-visit questionnaire**, 74% of the students answered correctly.

However, 19% of the students filling out the pre-visit questionnaire (and 6% in the post-visit questionnaire) answered that X-ray machines are always a source of radiation, even if they are switched off. Previous studies on students' conceptions suggest that *natural (cosmic & terrestrial) and human-made sources of ionising radiation are unknown to students* (Boyes & Stanisstreet, 1994; Colclough, 2007; Prather, 2000; Rego & Peralta, 2006; Riesch & Westphal, 1975). Therefore, a lack of knowledge about X-ray machines did not come as a surprise.

Moreover, 11% of the students answered in the pre-visit questionnaire (5% in the post-visit questionnaire) that the exposure to X-rays makes parts of the X-ray device radioactive. This answer is consistent with previous research results: students think that *irradiated material becomes radioactive* (Colclough, 2007; Henriksen & Jorde, 2001; Riesch & Westphal, 1975) and that *after an X-ray scan, objects and human bodies emit radiation* (Eijkelhof, 1990; Kaczmarek et al., 1987; Mubeen et al., 2008), see also subsection 2.4.2.

Also, 7% of the students in the pre-visit questionnaire (5% in the post-visit questionnaire) claimed that it would not be dangerous if the air had been exchanged before entering the room. Previous research results suggest that some students assume that *radiation consists of radiating or radioactive particles* (Millar & Gill, 1996; Neumann & Hopf, 2012; Prather, 2000), or students describe *ionising radiation as material* (Lijnse et al., 1990; Maidl & DeKay, 2012). Also, some students think that *radiation needs a medium such as air to spread* (Eijkelhof et al., 1990; Riesch & Westphal, 1975). All these conceptions would be

consistent with these students' answer choice in this item. Furthermore, some students probably assume that air would become radioactive when being exposed to X-radiation.

Last, 10% of the students in the pre-visit questionnaire (and 10% post) claimed that it would not be dangerous if the cleaning staff waits a while before entering. This is consistent with the conception of students that *radiation lingers for a while even without source* (Eijkelhof et al., 1990). Interestingly, this is the only wrong answer option that was chosen equally often by students before and after the S'Cool LAB hands-on session.

Tier-2 (confidence): Overall, students increased their confidence in their answers from 48% to 69% in the post-visit questionnaire.

Figure 32 displays a **Sankey diagram with the changes in students' answer behaviour from the pre- to the post-visit questionnaire**.

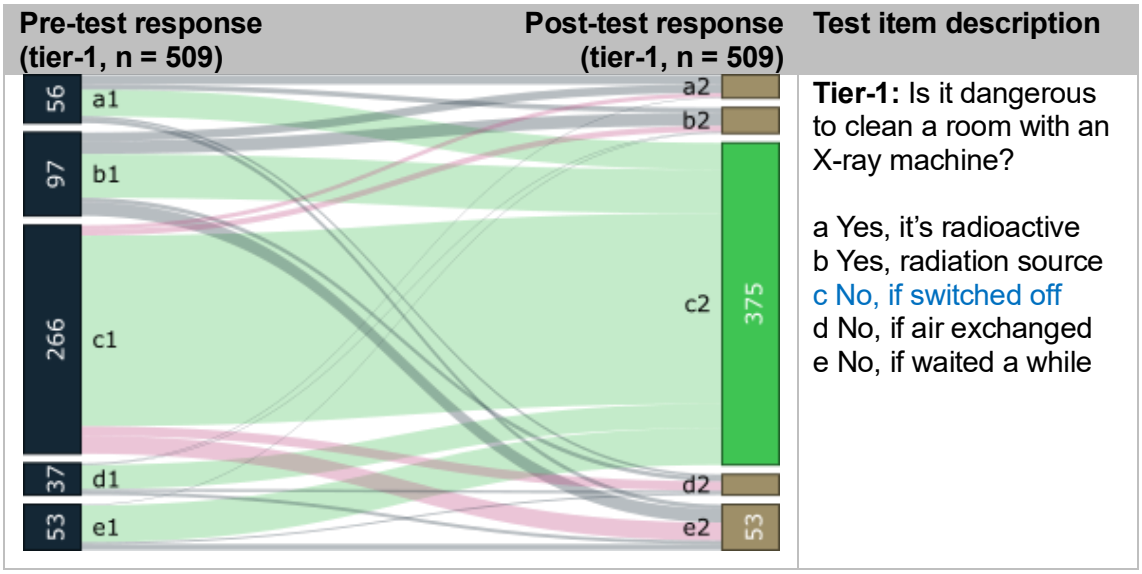


Figure 32. Sankey diagram: Dangers of cleaning a room with an X-ray machine (item XC). The diagram shows students' answers to the tier-1 level of the item before (pre-test, left side) and after (post-test, right side) taking part in the S'Cool LAB hands-on session.

Many students (30%) improved their answer behaviour at tier-1 from a wrong to the correct option. Only 7% of the students chose the same wrong option again after taking part in the S'Cool LAB hands-on session. Here, the most 'stable' wrong answer is option 2, stating that X-ray machines are always a source of radiation, even if they are switched off. 9% of the students changed their answer in an unfavourable way from the correct to a wrong option. Of those, the most frequent wrong answer was option 5, saying that it would not be dangerous it the cleaning staff waits a while before entering. Indeed, this wrong answer option is the most popular distractor among students filling out the post-visit questionnaire. One possible reason could be that participation in the X-ray workshop fosters conceptions that could lead students to choose this answer options. For example, when students

switched off the X-ray machines in S'Cool LAB, they always had to wait a few seconds before trying to open the doors because the automatic mechanical unlocking mechanism needed a few seconds to react. Although this was a purely mechanical problem, students might retrospectively assume that the waiting time was required to ensure radiation safety.

Item XW: A watch is exposed to X-radiation for 1 hour

Tier-1: In the **pre-visit questionnaire**, 37% of the students answered correctly that there is no danger associated with wearing a watch that had been X-rayed for 1 hour as long as no radioactive atoms were transferred to the watch. In the **post-visit questionnaire**, 66% of the students answered correctly.

The most popular distractor in the pre-visit questionnaire was option 1 ("*Yes, ionizing radiation accumulates in the watch, therefore the watch will emit radiation afterwards*."), which was chosen by 34% of the students. In the post-visit questionnaire, only 10% still chose this option. Indeed, previous research results suggest that certain students think that *radiation accumulates inside objects* (Alsop, 2001; Eijkelhof et al., 1990; Riesch & Westphal, 1975). Also, some students think that *radiation is conserved; if an object absorbs radiation it will later release the radiation* (Eijkelhof, 1990). Both conceptions would be consistent with this answer option.

Also, 16% of the students in the pre- (and only 6% in the post-) visit questionnaire said that since the watch had been exposed to X-rays, the atoms in the watch became radioactive. Again, this answer is consistent with previous research results on students' conceptions, see subsection 2.4.2.

Last, 13% of the students in the pre-visit questionnaire said that it would not be dangerous to wear the watch because the watch needs to be exposed to X-rays for a longer time. After the hands-on session, the proportion of students who answered in this way increased slightly to 18% in the post-visit questionnaire. Interestingly, this is the only wrong answer option that was chosen by more students after the hands-on session than before, which suggests a side-effect of the X-ray workshop. Indeed, this wrong answer option was the most popular distractor among students filling out the post-visit questionnaire. One possible reason could be that participation in the X-ray workshop fostered conceptions that could lead students to choose this answer option. In particular, students measured whether radiation was emitted by small bags of salt and pepper after they had been irradiated with X-radiation. However, students irradiated the material only for a few minutes in S'Cool LAB due to time constraints of the workshop. Here, students might retrospectively assume that the result of their measurement might have been different for a longer irradiation period.

Indeed, Prather (2000) found that some students think *the duration of radiation exposure determines how radioactive objects get*.

Tier-2 (confidence): Overall, students increased their confidence in their answers from 43% to 64% in the post-visit questionnaire.

Figure 33 displays a **Sankey diagram with the changes in students' answer behaviour from the pre- to the post-visit questionnaire**. Many students (39%) improved their answer behaviour at tier-1 from a wrong to the correct option. However, a small number of students (10%) changed their answer in an unfavourable way from the correct to a wrong option. Of those, the most frequent wrong answer was option 3, saying that the watch would need more exposure time to become radioactive itself.

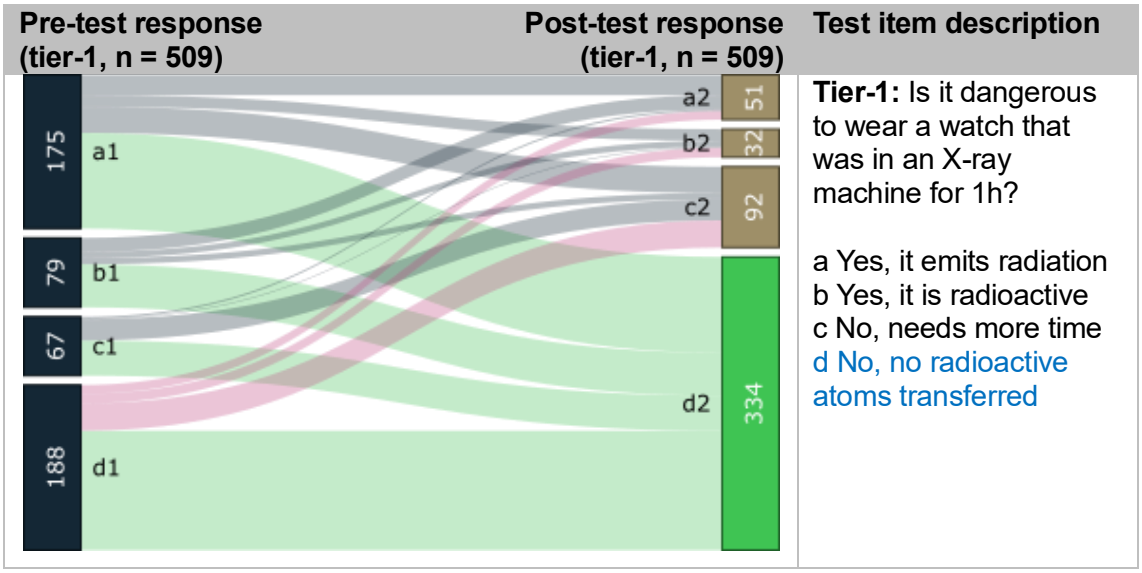


Figure 33. Sankey diagram: A watch Is exposed to X-radiation for 1 hour (item XW). The diagram shows students' answers to the tier-1 level of the item before and after taking part in the S'Cool LAB hands-on session.

7.4.4.4 Summary of items measuring conceptual learning

In summary, the seven chosen multi-tier multiple-choice items proved suitable to measure students' conceptual learning during the S'Cool LAB hands-on session, which consisted of three 90-minute hands-on workshops (cloud chamber workshop, electron tube workshop, and X-ray workshop). Also, almost all distractors generated from the results obtained within the pilot study (see section 6.4) were suitable by attracting at least 5% of the students. Furthermore, students' answers provided interesting insights into their reasoning patterns, which were consistent with previous research results on students' conceptions.

Electron tube workshop: the Lorentz force and protons in magnetic fields

In particular, students' misconceptions of the attractive forces of magnetic poles described by Maloney et al. (2001) were reproduced. Also, many answers were consistent with the belief that electrically charged particles are always affected by a magnetic field, even if they are at rest or move parallel to the field (Maloney et al., 2001). Scaife and Heckler (2011) summarised that these conceptions are very persistent even after instruction. Here, the positive effect of the short 90-minute electron tube workshop in S'Cool LAB is encouraging. It seems that the direct cognitive conflict with students' conception by letting them deflect electron beams with bar magnets from different angles can positively affect students' answer behaviour.

However, students' learning process was often not complete. In particular, the analysis of the two situations in which no force acts on the proton (items PR and PS) offered interesting insights into the side-effects of the hands-on session. Both of these special cases were not discussed directly during the electron tube workshop. Instead, they were included to estimate students' transfer skills towards unfamiliar situations. However, after students were introduced to the right-hand-rule, they readily applied this rule to all new situations and derived different directions of acting forces – even if the respective scenario involved a particle at rest or moving along the magnetic field lines. Although the positive effects of the hands-on session clearly outweighed these negative effects, future workshops should focus more on the differentiation between different cases. Relatively low confidence scores of approximately 50% in the post-questionnaire confirmed that students were not too sure about their answers.

Cloud chamber workshop: particle detectors on mountains and in space

A significant percentage of students answered according to the historic misconception that the *Earth is the main source of ionising particles*, see also 2.4.2. This conception persisted even after a 90-minute cloud chamber workshop, which included observing terrestrial and cosmic radiation. However, students' answers confirmed a by far more powerful distractor originating from the results of the pilot study (subsection 6.4.2). Indeed, many students expected a particle detector to detect air molecules instead of high-energy particles. Most likely, this can be explained by the undifferentiated use of the terms 'particle' and 'particle detector': Particle physicists use 'particle' most often for elementary particles in high-energy physics, and 'particle detector' for detectors able to measure high-energy particles (mostly ionising particles). However, in chemistry and physics school education, 'particle' is often used to describe atoms, molecules, or even dust particles. Hence, it is no surprise that students who are unfamiliar with high-energy physics expect 'particle detectors' to detect the 'particles' they know, such as air molecules. Consequently, it is not clear whether

this reasoning pattern should be interpreted as a misconception because it also reflects the difficulty of testing students' conceptions in a field of physics that involves many technical terms unknown to novices. In this case, the multiple-choice items might not measure conceptual understanding but rather the knowledge of physics vocabulary in the respective field.

Overall, the effect of the short 90-minute cloud chamber workshop in S'Cool LAB on students' answers to two multi-tier items was very positive. However, relatively low confidence scores of approximately 54% in the post-questionnaire (compared to 44% in the pre-questionnaire) suggest that students were not too sure about their answers.

X-ray workshop: cleaning an X-ray treatment room and X-raying a watch

Students' answers to the two items associated with the dangers of X-radiation confirmed students' conceptions described by previous studies. In particular, in the pre-visit questionnaire, students' answers were consistent with the belief that *irradiated material becomes radioactive* (Colclough, 2007; Henriksen & Jorde, 2001; Riesch & Westphal, 1975) and that *after an X-ray scan, objects and human bodies emit radiation* (Eijkelhof, 1990; Kaczmarek et al., 1987; Mubeen et al., 2008). However, the most powerful distractor was related to the conception that *radiation accumulates inside objects* (Alsop, 2001; Eijkelhof et al., 1990; Riesch & Westphal, 1975).

Overall, the effect of the short 90-minute X-ray workshop in S'Cool LAB on students' answers to these two items was very positive. Also, students increased their average confidence from 46% to over 67%. However, comparing students' answers between the pre- and post-visit questionnaire suggested a side-effect of the hands-on session. Indeed, the same number of students incorrectly answered that waiting a while makes it safe to enter an X-ray treatment room. Furthermore, more students incorrectly answered that it would require a longer irradiation time for stable atoms to become radioactive. These ideas were not discussed explicitly during the X-ray workshops in S'Cool LAB. Here, future workshops could integrate them into the experimentation and discussion phases.

7.4.5 Summary with respect to the research questions

In the following, the pre-post effects of the S'Cool LAB hands-on session are summarised with respect to the corresponding research questions. In addition, the results are compared with previous studies. However, as for some of these studies information was missing (e.g., missing standard deviations, missing variables pre or post, or missing correlations between pre- and post-test values), an effect size could only be estimated retrospectively for some of these studies. Furthermore, the absolute values of the scales were converted

to the percent of maximum possible score allow a direct comparison, as suggested by Cohen et al. (1999). Here, comparisons with similar out-of-school learning labs should be interpreted with caution because the sample taking part in S'Cool LAB workshops was biased towards students with high physics interest and achievement (see section 7.2).

► ***RQ1.1a) To which extent can half-day hands-on sessions in S'Cool LAB foster students' physics interest?***

Although the students taking part in the hands-on sessions in S'Cool LAB were already interested in physics beforehand (66% of the scale's maximum), they demonstrated even higher levels of situational interest with respect to the activities in S'Cool LAB (77% of the scale's maximum). Here, participation in half-day hands-on sessions in S'Cool LAB led to a **medium-sized effect on students' physics interest ($d_C = 0.6$)**, despite the relatively short intervention time of only 4.5 hours. Compared to similar studies on hands-on out-of-school learning labs using similar interest scales, S'Cool LAB fosters very high values of students' situational interest, which is based on a relatively high dispositional interest.

For example, Molz (2016) reported an average dispositional interest in physics of 45% of the scale's maximum for his treatment groups and a situational interest of 62% after hands-on physics learning activities (estimated average effect size over both treatment groups $d_C \approx 1.0$). In contrast, Hirth (2019) observed an opposite trend with a dispositional physics interest of 64% and a situational interest of 54% after hands-on learning activities on acoustics (estimated $d_C \approx -0.5$). However, due to the specific item wording in his study, the results for the situational interest are not easily comparable with the S'Cool LAB main study. Itzek-Greulich et al. (2017) reported a dispositional science interest of only 37% of the scale's maximum and situational interest of 56% of the scale's maximum after hands-on chemistry learning activities about starch. Only Pawek (2009) reported a similarly high situational interest for the students taking part in his study on hands-on physics learning activities corresponding to 70% of the scale's maximum. In a study with more than 11 000 students, Priemer et al. (2018) found an average value of 59% for the epistemic component of situational interest after taking part in hands-on out-of-school learning labs. **In summary, it seems that S'Cool LAB outperforms other hands-on out-of-school learning labs by triggering exceptionally high situational interest even for students whose dispositional interest in physics was already high.**

► ***RQ1.1b) To which extent can half-day hands-on sessions in S'Cool LAB foster students' curiosity towards particle physics?***

When comparing students' epistemic curiosity towards particle physics before and after taking part in hands-on S'Cool LAB workshops, **no significant difference** was identified.

However, the **level of students' curiosity towards particle physics was already very high in the pre-visit questionnaire (78% of the scale's maximum)**, which left little room for improvement towards the post-visit questionnaire (77% of the scale's maximum).

Indeed, previous studies assessing students' triggered epistemic curiosity with a similar scale and in a similar setting found significantly lower values for students triggered curiosity, for example, 48% for oscillations and waves (Hochberg, 2016), 62% for acoustics (Hirth, 2019), and 65% for radioactivity (Molz, 2016). However, these studies did not assess students' curiosity about these topics before the hands-on session. Therefore, no comparable values for a pre-post effect are available. **In summary, the results suggest that particle physics is an extremely promising topic to accommodate and maintain exceptionally high levels of epistemic curiosity – not just for out-of-school learning settings. However, hands-on S'Cool LAB workshops were not able to further increase students' curiosity towards particle physics.**

► ***RQ1.2a) To which extent can half-day hands-on sessions in S'Cool LAB foster students' physics self-beliefs?***

Similar to students' dispositional interest, their physics self-beliefs were already high before coming to S'Cool LAB (66% of the scale's maximum). Nevertheless, students reported even higher levels of situational self-beliefs with respect to the hands-on session in S'Cool LAB (74% of the scale's maximum). **Here, participation in half-day hands-on sessions in S'Cool LAB resulted in a small effect on students' physics self-beliefs ($\boldsymbol{d_C = 0.4}$).** Considering the relatively short intervention time of only 4.5 hours and considering that the improvement of students' self-beliefs was not the main focus of the hands-on session, this result is promising and consistent with previous findings. Indeed, a meta-analysis by O'Mara et al. (2006) reports a mean effect size of $d_C = 0.5$ for the impact of self-concept interventions, potentially of much longer duration.

Also, previous studies employing similar scales before and after similar hands-on sessions reported comparable or weaker results. For example, Molz (2016) reported almost the same trend with an average dispositional physics self-concept of 67% of the scale's maximum for his treatment groups and a situational physics self-concept of 75% of the scale's maximum (estimated average effect size over both treatment groups $d_C \approx 0.5$). In contrast, Pawek (2009) reported only a small effect from 67% for the dispositional to 70% for the situational self-concept ($d_C = 0.2$). Furthermore, Hirth (2019) described a slight decline in students' self-concept from 73% for the dispositional to 70% for the situational self-concept (estimated $d_C = -0.2$). However, due to the specific item wording in his study the results are not easily comparable with the S'Cool LAB main study.

In summary, it seems that S'Cool LAB outperforms other hands-on out-of-school learning labs by triggering very high situational physics self-beliefs even for students whose dispositional physics self-beliefs were already very high.

► ***RQ1.2b) To which extent can half-day hands-on sessions in S'Cool LAB foster students' conceptual understanding?***

In the pre-visit questionnaire, students demonstrated a low understanding of key concepts relevant to the hands-on workshops in S'Cool LAB with an average score of 39%. After taking part in half-day hands-on session, it was still not easy for them to answer the selected concept test items, but their average score improved to 56%. **Indeed, the most substantial effect of half-day hands-on sessions in S'Cool LAB was identified for students' conceptual understanding which showed a medium-sized increase from the pre- to the post-visit questionnaire ($d_C = 0.7$, Hake's gain[12] $g = 0.3$).**

Previous studies also confirmed positive learning gains of hands-on out-of-school learning labs. However, these studies mostly assessed gains in factual knowledge – not in conceptual understanding. For example, Huwer (2015) reported small to medium-sized effects on students' chemistry knowledge after participation in hands-on workshops (Cohen's $d_C = 0.3 - 0.5$ for different topics). Compared to learning new facts, conceptual learning requires special attention to students' pre-instructional conceptions. Although it seems that the regular school setting better supports learning and many out-of-school learning labs view learning rather as a 'side-effect', Molz (2016) included several concept test items in his achievement test. His results confirm that even a short intervention with hands-on learning activities in an out-of-school learning lab about radioactivity can reduce the percentage of students answering according to well-known misconceptions about irradiation and contamination. Overall, he also reported medium-sized learning gains (Hake's gain $g = 0.5$) with even slightly higher learning gains for the subscale representing the irradiation/contamination conceptions ($g = 0.6$).

In summary, half-day hands-on sessions in S'Cool LAB foster students' conceptual understanding significantly without compromising the impact on affective variables such as students' situational physics interest, or on students' physics self-beliefs.

[12] As discussed in subsection 5.4.2, Hake's gain is biased against low pre-test scores, whereas Cohen's d_C is better-suited for pre-post comparison because it takes the correlation between these two measures into account (Nissen et al., 2018). Here, Hake's gain is reported as an exception to facilitate the comparison of the study results with other studies.

► ***RQ3) To which extent do female and male participants differ in their educational outcomes?***

No gender difference was found when comparing students' curiosity towards particle physics or their conceptual understanding measured during the pre- and post-visit questionnaires. Although girls reported a slightly lower dispositional interest in physics ($d_C = 0.3$) and lower dispositional physics self-beliefs ($d_C = 0.4$), no gender differences were identified for students' situational interest and self-beliefs. Therefore, with respect to interest (IE) and self-beliefs (SC), participation in the S'Cool LAB hands-on session had a stronger effect on girls ($d_{C,IE} = 0.8, d_{C,SC} = 0.5$) than on boys ($d_{C,IE} = 0.5, d_{C,SC} = 0.3$). It seems that girls, in particular, benefit from the hands-on sessions, which is consistent with previous findings, for example, by Burkam et al. (1997) or Engeln (2004). While these effects on the 'gender gap' may not appear large, they refer to rather stable traits. Hence, it is remarkable that even a short half-day intervention can initiate a positive change. Similarly, Pawek (2009) noticed an even larger gender gap in students' dispositional self-beliefs ($d_C = 0.9$) and interest in physics ($d_C = 0.9$). Moreover, he reported smaller gender differences in students' situational interest after taking part in a hands-on out-of-school learning lab (average over three interest components $d_C = 0.3$). In a large study with more than 11 000 students, Priemer et al. (2018) report a negligible gender difference in students' situational interest after taking part in hands-on activities at out-of-school learning labs (estimated Cohen's $d_C \approx 0.1$). **In summary, female and male participants do not differ in their educational outcomes after taking part in half-day hands-on sessions in S'Cool LAB.**

7.5 Correlations among S'Cool LAB's educational outcomes

Most of the eight educational outcomes assessed in the framework of this study are closely related and intertwined theoretical constructs, which is reflected in the concept map presented in subsection 2.2.7. For example, helpfulness, enjoyment, and cognitive activation are closely related to interest or interest development. Also, students' perception of the easiness of the activities in S'Cool LAB may contribute to their situational self-beliefs. However, it is not always clear if variables are only correlated, or whether certain variables are caused by other variables. For example, Förtsch et al. (2017) describe the classroom criterion cognitive activation as predictor of situational interest while Lipowsky (2009) describes cognitive activation as mediator between interest and learning outcomes.

Furthermore, all variables considered as educational outcomes in this study were measured at the same time approximately 30 days after the hands-on session in

S'Cool LAB. Therefore, there was no temporal distinction between the variables that could be used to establish cause-effect relationships through, for example, structural equation modelling. Consequently, these eight variables were modelled as separate educational outcomes and their correlations among each other were studied separately without attempting to model cause-effect relationships. Nevertheless, these correlations provided interesting information about the intertwined nature of various educational outcomes of hands-on sessions in S'Cool LAB. Here, Table 71 displays the correlation matrix of all eight educational outcomes.

Furthermore, Figure 34 compares the concept map of the underlying theoretical constructs of these educational outcomes (see subsection 2.2.7) with a map of high correlations that were calculated based on the data of the S'Cool LAB main study. In general, almost all dependent variables were significantly correlated, most of them with medium to high correlation coefficients. This is consistent with the results reported by Cors (2016), who summarised that "*correlations among dependent variables* [science and technology interest, attitude, self-concept and career aspirations] *were virtually all significant and most are medium, with some variation:* $0.12 \leq r \leq 0.76$".

Table 71. Pearson correlation r between the eight educational outcomes and associated p-values. Significant correlations (Bonferroni correction $m = 28, p < 0.002$) are highlighted: high correlations $r > 0.5$ in green, moderate correlations $r = 0.2 - 0.5$ in yellow, small correlations $r < 0.2$ in blue.

		IE3	CS3	ENJ	HEL	SC3	EAS	CA	UN3
Situational interest IE3	r	1	**0.77**	**0.66**	**0.66**	**0.54**	**0.44**	**0.65**	0.09
	p		2E-101	5E-64	1E-64	2E-40	6E-25	1E-61	0.067
Curiosity state particle physics CS3	r		1	**0.54**	**0.56**	**0.42**	**0.32**	**0.58**	**0.15**
	p			2E-39	2E-43	1E-23	3E-13	4E-46	0.002
Perceived enjoyment ENJ	r			1	**0.64**	**0.45**	**0.49**	**0.62**	0.09
	p				4E-61	2E-26	1E-31	8E-56	0.05
Perceived helpfulness HEL	r				1	**0.36**	**0.32**	**0.61**	-0.06
	p					3E-17	7E-14	1E-52	0.217
Situational self-beliefs SC3	r					1	**0.55**	**0.57**	0.14
	p						4E-42	3E-44	0.003
Perceived easiness EAS	r						1	**0.45**	**0.18**
	p							1E-26	1E-04
Perceived cognitive activation CA	r							1	0.12
	p								0.009

In particular, students' situational interest and self-beliefs were strongly correlated with each other ($r_{IE3,SC3} = 0.54, p = 2 \cdot 10^{-40}$). This was expected since both can be interpreted as components of motivation. Here, Hirth (2019) reported a slightly higher correlation ($r = 0.72,\ p < 0.001$) whereas Itzek-Greulich et al. (2017) reported a slightly lower correlation ($r = 0.45, p < 0.001$) for effects of a hands-on sessions in out-of-school learning labs. Also, perceived easiness was correlated with students' situational self-

beliefs. This comes as no surprise since self-beliefs are usually updated based on the perception of one's performance (see subsection 2.2.2).

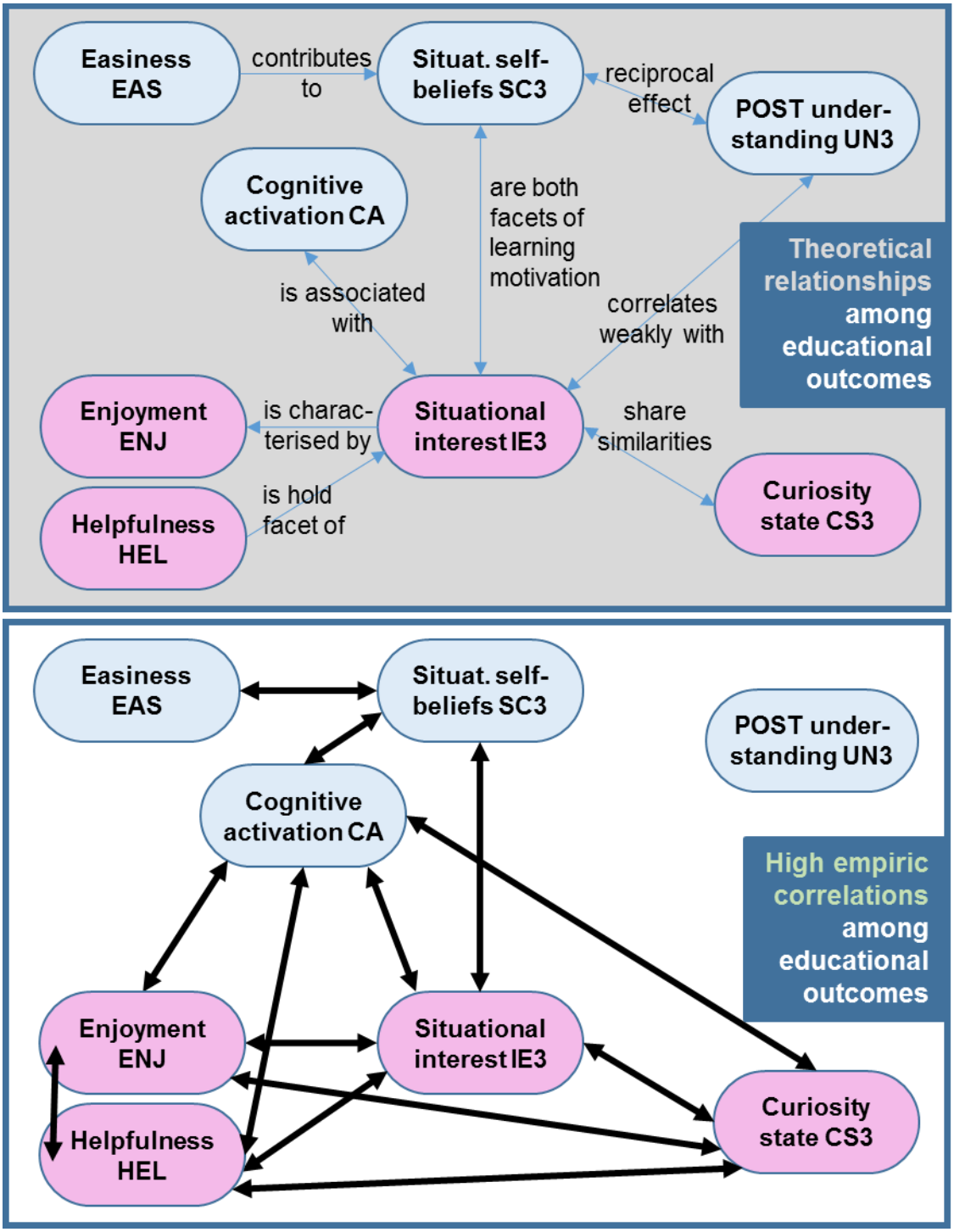


Figure 34. Above: Concept map of educational outcomes as presented in subsection 2.2.7. Below: High and significant correlations between educational S'Cool LAB outcomes, connection lines indicate correlations coefficients $r \geq 0.5, p < 0.002$.

When focusing on the more affective outcomes (displayed in pink in Figure 34), the data show all of them to be strongly correlated with each other and with students' perceived cognitive activation. It is not surprising that enjoyment, perceived helpfulness, and curiosity

were correlated with situational interest and, consequently, with each other. In particular, enjoyment can be interpreted as characteristic of interest. Here, Itzek-Greulich et al. (2017) reported an even higher correlation for these two outcomes of hands-on activities ($r = 0.79, p < 0.001$).

Moreover, cognitive activation is also associated with interest while helpfulness can be interpreted as 'hold' component in interest development. Finally, epistemic curiosity and interest are very similar constructs. Specifically, cognitive activation seems to link the more affective with the more cognitive outcomes as it was strongly correlated with all affective outcomes but also with situational self-beliefs and, to a slightly lesser degree, with perceived easiness.

Students' conceptual understanding after taking part in a hands-on session in S'Cool LAB was only very weakly correlated with two other educational outcomes, namely with their curiosity towards particle physics ($r_{CS3,UN3} = 0.15, p = 0.002$) and their perception of easiness ($r_{EAS,UN3} = 0.18, p = 0.0001$). Here, previous studies reported slightly higher correlations between affective and cognitive outcomes. For example, Molz (2016) reported a correlation of $r = 0.3$ between achievement and motivation. However, based on a meta-analysis, Hansford and Hattie (1982) reported an average correlation between self-measures (such as self-concept or self-esteem) and achievement of $r = 0.21$, which confirms a similar value.

Overall, correlations between S'Cool LAB's educational outcomes were consistent with the theoretical foundation of these variables as described in section 2.2, and their intertwined nature was confirmed.

7.6 Additional questions as part of the formative assessment of S'Cool LAB

Besides S'Cool LAB's educational outcomes and novelty factors, this study also aimed to provide information for CERN and the S'Cool LAB team to be used in the framework of the formative assessment of this out-of-school learning opportunity. These formative assessment questions were introduced in section 3.3. Here, subsection 7.6.1 compares students' perception of the S'Cool LAB hands-on session and their perception of the guided tour at CERN. Subsection 7.6.2 compares how students perceive the three workshops in S'Cool LAB with respect to enjoyment and cognitive load. Finally, subsection 7.6.3 describes how important the different aspects of the trip to CERN were for students.

7.6.1 Perception of the hands-on session compared to the guided tour at CERN

As part of the post-visit questionnaire, students were asked to rate their perceived satisfaction, the perceived duration and their perceived amount of learning for the guided tour at CERN (2.5 hours), the hands-on session in S'Cool LAB (4.5 hours), and the whole day at CERN including S'Cool LAB (8.5 hours, including breaks), see Table 72. These items were then analysed to answer the question

- ***Q1) How do students perceive the hands-on session in S'Cool LAB compared to their guided tour at CERN?***

In general, students enjoyed their day at CERN very much with an average value of 86% of the scale's maximum. Moreover, they liked S'Cool LAB significantly more than their guided tour ($d_C = 0.4, 95\%\ CI\ [0.3, 0.6]$) and they also reported having learned more in S'Cool LAB than during the guided tour ($d_C = 0.4, 95\%\ CI\ [0.2, 0.5]$).

However, the guided tour was much shorter than the hands-on session in S'Cool LAB and, thus, provided less time to develop educational outcomes. Moreover, due to different activity types and content, the comparability of guided tours and hands-on sessions in S'Cool LAB may be limited. Nevertheless, these results support the additional value of hands-on experimentation as part of a visit to CERN.

Table 72. Students' perception of the S'Cool LAB hands-on session compared to their guided tour at CERN. Students' satisfaction, the perceived duration and their perceived learning is reported through the mean value (M) of a seven-level scale (1-7) and the respective standard deviation (SD). In addition, the mean value in percent of maximum possible score (M%) is reported.

How did you like your S'Cool LAB day at CERN? (1 not at all – 7 very much)	**M**	**SD**	**M%**
The guided tours at CERN	5.81	1.28	80%
Working on the experiments in S'Cool LAB	6.31	0.93	89%
The whole day at CERN including S'Cool LAB	6.18	0.96	86%
How do you feel about the durations? (1 too long – 7 too short)			
The guided tours at CERN	3.86	1.28	48%
Working on the experiments in S'Cool LAB	3.97	1.30	50%
The whole day to CERN including S'Cool LAB	3.68	1.29	45%
How much did you learn? (1 nothing at all – 7 very much)			
During the guided tours at CERN	5.41	1.23	73%
While working on the experiments in S'Cool LAB	5.86	1.21	81%
During the whole day at CERN including S'Cool LAB	5.86	1.04	81%

When asked about the duration, students perceived both the guided tour and the S'Cool LAB hands-on session as neither too long nor too short. However, S'Cool LAB tutors often reported that students were very tired at the end of the day. At the same time, students

had to rush through some of the experimental tasks to finish the workshops in time. Therefore, it was decided to reduce the number of experiments from three to two while increasing the amount of time per workshop slightly for future programmes in S'Cool LAB. These changes were implemented after the data collection for the S'Cool LAB main study was completed.

7.6.2 Perception of the S'Cool LAB workshops

- ***Q2) Which of the workshops in S'Cool LAB do students like most?***
- ***Q3) Which of the workshop is associated with a high amount of cognitive load?***

To answer these two questions, students were asked to answer how much they liked each of the three workshops in S'Cool LAB on a seven-level scale from not at all (coded as 1) to very much (coded as 7) after taking part in the hands-on session in S'Cool LAB. Here, a seven-level scale was chosen to keep the item format similar to the other questions in the respective item block, which are reported in Table 72. Later in the questionnaire, students were also asked to report their perceived cognitive load for every workshop by rating four statements on a six-level scale form disagree completely (coded as 1) to agree completely (coded as 6). The detailed item texts are included in the post-visit questionnaire in appendix 10.5. For details about the item properties of the cognitive load scales, see subsection 6.3.5.

Table 73 summarises students' perception of the different S'Cool LAB workshops. Overall, students liked all workshops and reported low values for their perceived cognitive load. Therefore, the development of the different learning activities can be rated successful as it led to enjoyable experiences while providing enough support for students to handle the high-tech equipment. While the electron tube and X-ray workshop were perceived as very similar with respect to cognitive load and liking, the cloud chamber workshop was the clear favourite of the students reaching, on average, 92% of the scale's maximum.

However, contrary to the other workshops, students did not follow and fill out worksheets during the cloud chamber workshop mostly because of the light conditions (the observation of the cloud chambers required darkness). Instead, the tutor demonstrated how to assemble the detector and discussed the observations together with all students in front of a whiteboard after they had the chance to make observations with their detector. This different format could explain the difference to the other two workshops. However, students might also prefer the cloud chamber workshop because it includes captivating visual effects from natural radiation as well as working with dry ice.

Students of different age or gender might perceive the workshops differently. Therefore, independent t-tests were calculated to compare the liking and cognitive load scores for students below and above the mean age of 17.0 years and both for girls and boys. Here no significant age differences were detected.

However, girls liked the cloud chamber workshop slightly more than boys ($t(464) = -3.8$, $p = 0.0002, d_C = 0.3, 95\%\ CI\ [0.1,0.5]$). Contrary, girls liked the electron tube workshop slightly less than boys ($t(502) = 3.1, p = 0.002, d_C = 0.3, 95\%\ CI\ [0.1,0.5]$), which was associated with a slightly higher perceived cognitive load for this workshop ($t(338) = -2.1$, $p = 0.03\ d_C = 0.2, 95\%\ CI\ [0.0,0.4]$).

Table 73. Students' perception of the different S'Cool LAB workshops. Students' liking of the workshops is reported through the mean value (M) of a seven-level scale (1-7) and the respective standard deviation (SD). In addition, the mean value in percent of maximum possible score (M%) is reported. Similarly, students' perceived cognitive load is reported with respect to a six-level scale. Scale values are reported in total and separately for female and male participants.

	Liking 7-level items			Cognitive load 6-level items		
Experiment	**M**	**SD**	**M%**	**M**	**SD**	**M%**
Cloud Chamber	**6.47**	**0.91**	**92%**	**1.94**	**1.05**	**19%**
Male participants ($n = 325$)	**6.37**	0.99		1.98	0.82	
Female participants ($n = 178$)	**6.66**	0.71		1.87	0.79	
Experiments with the electron tube	**5.56**	**1.33**	**77%**	**2.26**	**1.21**	**25%**
Male participants ($n = 325$)	**5.69**	1.28		**2.19**	0.97	
Female participants ($n = 178$)	**5.31**	1.39		**2.39**	1.05	
Experiments with the X-ray machine	**5.59**	**1.34**	**77%**	**2.15**	**1.18**	**23%**
Male participants ($n = 325$)	5.55	1.37		2.11	0.95	
Female participants ($n = 178$)	5.66	1.29		2.22	1.01	

Overall, these results confirmed the successful development of the workshops in S'Cool LAB. In particular, students like the different workshops independent of their age and mostly independent of their gender. However, since students reported low values for their perceived cognitive load, more advanced-level workshops were developed after the S'Cool LAB main study to challenge, in particular, more experienced students. Here, the superconductivity workshop was developed to allow students to compare the electrical resistance of a normal conductor with a high-temperature superconductor (Bi-2223) and to study the Meissner-Ochsenfeld-Effect and the Flux-Pinning-Effect. Furthermore, the Positron-Emission-Tomography workshop allows students to calibrate and use scintillation detectors to understand the basic principles of Positron-Emission-Tomography (PET) and to locate a positron source (Na-22).

7.6.3 Important aspects of students' trip to CERN

► ***Q4) How do students rate the importance of different aspects of their trip to CERN?***

To answer this question, the post-visit questionnaire included seven items based on Pawek (2009) that had to be rated on a six-level scale from 'not at all important' (coded as 1) to 'very important (coded as 6). Pawek (2009) found that, in his follow-up questionnaire, students reported 'fun' as the most important aspect of their visit to an out-of-school learning lab in Germany followed by 'doing experiments', 'learning', and lastly, 'no school'. After the S'Cool LAB Day programme, it was expected that the 'CERN halo' overshadowed the experiences during the hands-on session in S'Cool LAB. Here, Table 74 summarises students' perception of the different aspects of their trip to CERN.

Table 74. Students' perception of the different aspects of their trip to CERN is reported through the mean value (M) of a six-level scale from 'not at all important' (1) to 'very important (6) and the respective standard deviation (SD). In addition, the mean value in percent of maximum possible score (M%) is reported.

Item label and text		M	SD	M%
I7	Having visited CERN	5.53	0.81	91%
I2	Fun	5.27	0.89	85%
I5	Learning more about particle physics	5.17	0.88	83%
I4	Learning more about physics	5.14	0.88	83%
I6	Having visited S'Cool LAB	4.87	1.06	77%
I3	No school	3.44	1.58	49%

Indeed, as the most important aspect of their trip, students listed 'having visited CERN' with a mean value of 91% of the scale's maximum. However, more than 80% of the scale maximum is reached for 'fun', 'learning more about physics' as well as 'learning more about particle physics'. The least important aspect was that a trip to CERN involved 'no school' (49%). In summary, it comes as no surprise that visiting CERN was very important to students. Learning more about (particle) physics in a fun way was another important aspect of their trip. Doing experiments on their own and visiting S'Cool LAB were rated almost equally important. This suggests that the hands-on sessions in S'Cool LAB were able to stand out from the general 'CERN halo'.

7.7 S'Cool LAB's educational outcomes: sequential linear models (SLMs)

Based on novelty theory and cognitive load theory (subsections 2.3.3 and 2.3.2), learners' perceived novelty can affect the educational outcomes of out-of-school learning opportunities. High-quality learning activities need to provide a certain amount of novelty to trigger students' attention and interest while preventing cognitive overload through structure and organisation. Furthermore, certain student characteristics and previous

experiences can influence how students perceive novelty (novelty influence and experience factors, see 2.3.4).

Here, sequential linear regression models were calculated to estimate the importance of the different novelty influence factors and novelty experience factors with respect to S'Cool LAB's eight educational outcomes (see 7.3.14). Here, the different novelty factors were entered into a sequential linear model with decreasing expected predictive power as suggested by Urban and Mayerl (2006). Also, the assumed time of causation was considered when defining the order of predictors: 'older' and more stable variables, such as gender, age or dispositions entered the model earlier than 'newer' and less stable variables, such as personality states.

Consequently, novelty influence factors (NIFs) were entered before considering novelty experience factors (NEFs) because NEFs seem to mediate the effect of NIFs on educational outcomes (Cors et al., 2017) and because the assumed time of causation as well as the time of measurement was earlier for novelty influence factors compared to novelty experience factors. Within the group of NIFs, dispositions and stable character traits, such as interest in physics, self-beliefs or curiosity trait were assumed to be strong predictors of the corresponding situational variables, such as situational interest, self-beliefs or curiosity state. Hence, when modelling these situational variables, the corresponding disposition was entered into the regression model first or right after the control variables age and gender. Corresponding dispositions were then followed by closely related variables. Last, NEFs were considered after including all NIFs.

The following subsections include the sequential linear models on students' situational interest (7.7.1), curiosity state towards particle physics (7.7.2), perceived enjoyment and helpfulness (0), situational physics self-beliefs (7.7.4), perceived easiness (7.7.5), cognitive activation (7.7.6), and students' conceptual understanding (7.7.7) after taking part in the S'Cool LAB hands-on session. Here, the subsection on students' situational interest (7.7.1) is described in detail and includes, for example, all tests of the assumptions of linear regressions. All other educational outcomes were treated in an analogous way. However, the tests of assumptions are reported in a shortened way and details are included in appendix 10.9 instead. Subsection 7.7.8 provides a summary of all models.

7.7.1 SLM: Situational interest

A sequential regression was calculated to determine the novelty factors (NIFs & NEFs), which predict students' situational interest after taking part in the S'Cool LAB hands-on session. Predictors were selected based on their expected theoretical relationship to the

dependent variable and entered into the model with decreasing theoretical importance. Table 75 shows the order of predictors and includes a brief justification for each predictor.

Table 75. Order of predictors for the sequential regression on students' situational interest.

#	Variable	Label	Justification
1	Gender (femaleness)	SEX	Girls tend to report lower values for physics interest (Hoffmann et al., 1997)
2	Age	AGE	Interest declines with age (Todt & Schreiber, 1998; Wigfield & Eccles, 1992)
	Novelty Influence Factors	NIFs	Prior experiences and dispositions might affect students' perceived novelty and their educational outcomes (Cors et al., 2017)
3	Dispositional physics interest	IE1	Dispositional interest is an important predictor of situational components of interest (Krapp, 2002)
4	Curiosity state particle physics	CS1	Epistemic curiosity is very similar to the construct of interest (Alexander & Grossnickle, 2016)
5	Experience hands-on experiments	EH	NIFs (dimension 'capabilities'): directly related to the hands-on activities in S'Cool LAB with student worksheets in English
6	English skills (binary)	EN	
7	Experience with OSLEOs[13]	EO	NIF (dimension 'setting familiarity'): has been shown to predict educational outcomes (Cors, 2016)
8	Experience with particle physics	EP	NIF (dimension 'cognitive') suggested by (Orion & Hofstein, 1994)
	Novelty Experience Factors	NEFs	Students novelty experience at an out-of-school learning opportunity can affect educational outcomes (Cors et al., 2017)
9	Tinkering support (novelty)	TIN	NEF (dimension 'capabilities'): a similar construct (tendency to tinker) has been shown to predict educational outcomes (Cors, 2016)
10	Setting orientation (novelty)	SO	NEF (dimension 'setting familiarity'): relevant according to Cors et al. (2017) and Molz (2016)
11	Lack of problems (novelty)	PRO	NEF (dimension 'capabilities'): these factors are relatively similar, high cognitive load can hinder learning and development (Sweller, 1994)
12	Cognitive load experiments	CLE	
13	Tutor support	TS	NEF (dimension 'social'): this factor was identified as important predictor by previous studies (Orion & Hofstein, 1994; Pawek, 2009)

Evaluation of assumptions of sequential linear regression

The evaluation of the assumptions of linear modelling (compare subsection 5.5.2) is summarised in Table 76. Indeed, univariate analysis already identified violations of the normality assumptions for several of the included variables (see 7.3.14). In particular, many variables were negatively skewed indicating a ceiling effect, which can be explained by the special sample of high-achieving and motivated students (see 7.2). To evaluate whether the problematic properties of the individual variables are reflected also in the residuals of

[13] **O**ut-of-**s**chool **le**arning **o**pportunities

the linear models, the residual scatterplot (standardised regression residuals over standardised predicted regression values) was inspected, see Figure 35.

Table 76. Assumptions of sequential linear regression and results of their evaluation.

Assumptions	Target (see 5.5.2)	Main study result	
Ratio of cases to independent variables	$N \geq 50 + 8m$ $N \geq 104 + m$	$m = 13, N = 509$ $N \geq 154$	✓
Absence of multivariate outliers	Mahalanobis distance < 35 $(df = 13, \alpha = 0.001)$	14 critical cases[14]	↯
Normality	Rectangular scatterplot: z(regression residuals) over z(predicted regression values)	Ceiling effect for high z-values	↯
Homoscedasticity			↯
Linearity		OK	✓
Independent residuals	Durbin Watson $1.5 < DW < 2.5$	1.7	✓
Outliers in the solution	Standard. residuals $\lvert z \rvert \leq 3.29$	3 critical cases[15]	↯
Multicollinearity	correlations < 0.70	< 0.65	✓
Multicollinearity	tolerance > 0.1	> 0.5	✓

If the assumptions of linearity, homoscedasticity and normality were fulfilled, the residuals would display a rectangular distribution (without curvature) with a concentration of data points along the centre, a horizontal line at zero, and a normal distribution of residuals when going symmetrically away from this centre (Tabachnick & Fidell, 2013). In Figure 35, one can see a favourable distribution for a large range of predicted values. Also, the linearity assumption was met as there is no discernible curvature. However, in particular for high predicted z-values on the right, there is an asymmetrical clustering of data points for positive residuals, see red rectangle in Figure 35. Thus, the distribution of residuals is also skewed showing a deviation from a normal distribution, and violation of homoscedasticity. Furthermore, the scatterplot shows three outliers with extreme standardised regression residuals $|z| > 3.29$. Nevertheless, variables were not transformed to maintain comparability between the different predictors. Therefore, the standard errors of the predictors might be underestimated and a smaller alpha level may be appropriate when evaluating significant predictors by gauging their p-values.

Furthermore, the Mahalanobis distance was computed for all cases and compared to the critical value of $\chi_c^2 = 35$ ($p < 0.001$, $df = 13$). Consequently, 14 cases were identified as multivariate outliers, two of which were already identified as outliers in the solution due to the extreme z-values of their standardised residuals. Also, all 16 outliers showed extreme z-values for at least one of the predictor variables during univariate analysis (see appendix 10.8). Since no cases had missing data, the initial sample size ($N = 509$) was large

[14] Multivariate outliers: cases 72, 133, 146, 157, 187, 253, 269, 320, 328, 365, 378, 410, 434, 439
[15] Outliers in the solution with extreme residuals: cases 146, 305, 383

enough for the number of predictors, even after removing the 16 outliers. Also, the Durbin Watson statistic of 1.7 confirmed the independence of residuals.

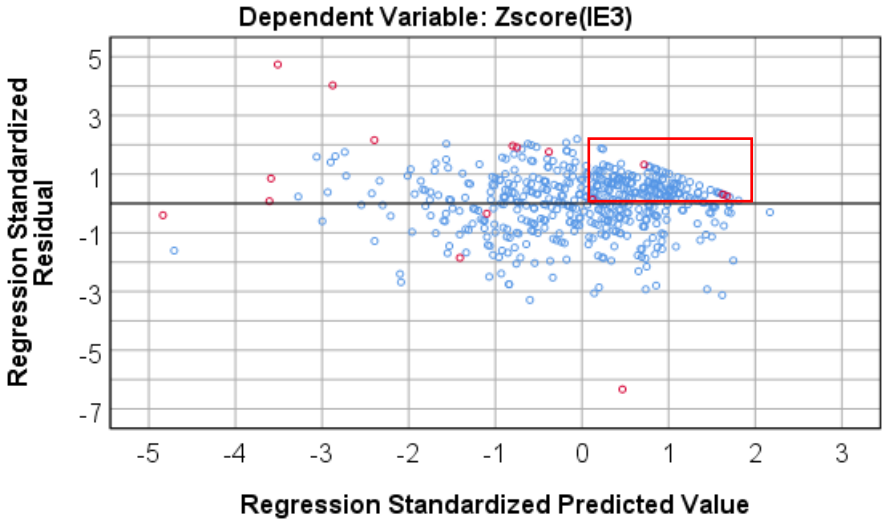


Figure 35. Residuals scatterplot for dependent variable situational interest (IE3). 16 cases displayed in red were indentified as multivariate outliers (14 cases) or outliers in the solution with extreme residuals (2 cases) and were excluded from the analysis. The red rectangle marks homoscedasticity issues.

In summary, the normality and homoscedasticity assumptions of the regression residuals were violated, even after removing outliers. Consequently, the standard errors of the predictors might be underestimated. However, the Bonferroni correction with $m = 13$ was already applied to take the high number of 13 predictors into account and prevent overfitting. This quite conservative correction reduces the critical alpha level significantly. Indeed, less restrictive corrections such as the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) are often used instead. Here, the more restrictive correction was chosen deliberately because it will also treat to a certain degree the underestimation of standard errors cause by the normality and homoscedasticity violations.

Appendix 10.9.1 lists the full correlation matrix for all predictors and the dependent variable. High correlations ($r > 0.5$) were identified between 'dispositional interest' and 'curiosity state particle physics' ($r_{IE1,CS1} = 0.65,\ p = 10^{-63}$), the novelty factor 'lack of problems' and 'cognitive load' ($r_{PRO,CLE} = -0.59,\ p = 10^{-49}$), and the novelty factor 'tinkering support' and 'tutor support' ($r_{TIN,TS} = 0.57,\ p = 3 \cdot 10^{-44}$). Indeed, interest and epistemic curiosity are very similar constructs (Alexander & Grossnickle, 2016), therefore, a high correlation between both constructs came as no surprise. Also, 'lack of problems' and 'cognitive load' are two scales with very similar items, although cognitive load items measured in the opposite direction (*"It was hard …"* versus *"It was easy… "*). Finally, 'tinkering support' and 'tutor support' both refer to students' perceived support while doing experiments.

Further significant but smaller correlations were identified among all of the novelty experience factors. Here, Cors (2016) reported significant and comparable correlations between all novelty experiences factors in her studies. Moreover, all novelty experience factors were correlated with the dependent variable 'situational interest'. However, all correlations were below the critical value of 0.70 and tolerance values > 0.5 confirmed that there were no critical multicollinearity issues for the data set. Nevertheless, the order in which these highly correlated predictors entered the linear model might have affected the size of their regression coefficients due to their shared variance. In particular, the regression coefficients for the different novelty experience factors might be difficult to interpret.

Results and discussion

Table 77 displays the standardised regression coefficients and p-values for full sample (β_{all}, p_{all}), and for the reduced sample without outliers $(\beta_{w/o}, p_{w/o})$ together with the 95% confidence interval for the standardised regression coefficients, and the adjusted R^2 per regression level. Indeed, predictors were entered step by step to examine, which predictors improved the model significantly and which regression level accounted for which proportion of variance of the dependent variable.

As expected, removing 16 outliers had only a small effect on the size of the regression coefficients and reduced, overall, their standard errors as reflected in lower p-values. Therefore, only the solution without outliers is interpreted below. With all 13 independent variables, the model accounted for a significant amount of 55% of the variance in students' situational interest $(Adjusted\ R^2 = 0.55, SEE = 0.64, F(13{,}479) = 41, p = 2 \cdot 10^{-77})$, which indicates a good model fit for the data according to Muijs (2004).

While the control variables age and gender did not account for any of the variance in the dependent variable, adding 'dispositional interest' to the model accounted for 26% of the variance of the situational interest. Adding students' curiosity state towards particle physics increased the adjusted R^2 significantly to 31%. However, the model did not improve when adding students' prior experiences (with hands-on experiments, OSLEOs, and particle physics) and their (dichotomised) English skills. Therefore, all novelty influence factors together accounted for almost one third (31%) of the variance in the dependent variable 'situational interest'. However, even after controlling for novelty influence factors, adding novelty experience factors improved the prediction model. Among the novelty experience factors, 'tinkering support' increased the adjusted R^2 significantly to 50% while students' perceived support by tutors added another 5%. When modelling students' situational epistemic interest after participating in a hands-on out-of-school learning lab, Pawek

(2009) reported similar results with 53% modelled variance in total. While 32% of the variance were modelled by considering only dispositions and personality traits (similar to novelty influence factors), 40% of the variance were modelled by considering only laboratory characteristics (which are similar to novelty experience factors).

In total, only four of the 13 regression coefficients differed significantly from zero. The significant predictors among the novelty influence factors were students' dispositional interest ($\beta_{w/o} = 0.21^{***}$) and curiosity towards particle physics ($\beta_{w/o} = 0.20^{***}$). Furthermore, the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.29^{***}$), and 'tutor support' ($\beta_{w/o} = 0.27^{***}$) improved the prediction significantly.

Table 77. Sequential regression of novelty influence and experience factors on students' situational interest: standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[16] and significant changes in the adjusted R^2 ($p < 0.004$)[17] are highlighted.

	full sample		w/o outliers			95% CI		
	$N = 509$		$N = 493$			Lower	Upper	Adj.
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)		0.405			0.601	-0.82	1.42	
Gender	0.02	0.538	0.03		0.274	-0.05	0.19	0.00
Age	-0.03	0.339	-0.02		0.495	-0.09	0.04	0.00
Dispositional physics interest IE1	**0.21**	2E-06	**0.21**	***	8E-07	0.13	0.29	**0.26**
Curiosity state particle phys. CS1	**0.18**	6E-05	**0.20**	***	2E-06	0.12	0.29	**0.31**
Experience hands-on experim. EH	-0.06	0.058	-0.03		0.277	-0.09	0.03	0.31
English skills (binary) EN	0.03	0.347	0.04		0.185	-0.04	0.20	0.31
Experience with OSLEOs EO	-0.01	0.660	0.01		0.777	-0.07	0.09	0.31
Experience with particle physics EP	-0.04	0.280	-0.04		0.265	-0.09	0.03	0.31
Tinkering support (novelty) TIN	**0.23**	7E-07	**0.29**	***	3E-10	0.20	0.37	**0.50**
Setting orientation (novelty) SO	0.10	0.012	0.07		0.074	-0.01	0.14	0.51
Lack of problems (novelty) PRO	-0.07	0.094	-0.07		0.077	-0.14	0.01	0.51
Cognitive load experiments CLE	-0.07	0.110	-0.04		0.321	-0.12	0.04	0.51
Tutor support TS	**0.27**	6E-11	**0.27**	***	3E-11	0.20	0.37	**0.55**
Full model:	$Adjusted\ R^2 = 0.52$		$Adjusted\ R^2 = 0.55, SEE = 0.64$ $F(13,480) = 47, p = 2 \cdot 10^{-77}$					

These results suggest that higher levels of situational interest were reached by students who were already interested in physics and curious about particle physics. Furthermore,

[16] The Bonferroni correction reduces the required p-values with $m = 13$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

[17] This $p < 0.004$ threshold is applied to gauge the significance of a change in the F-values that correspond to the respective adjusted R^2 of two consecutive regression level. As the predictors are entered step by step, the number of regression levels equals the number of predictors. Here, the initial alpha level of 0.05 is adjusted using the Bonferroni correction.

students who perceived the learning atmosphere as favourable with high levels of tinkering and tutor support benefited more from the hands-on session in S'Cool LAB with respect to the educational outcome 'situational interest'.

7.7.2 SLM: Curiosity state particle physics

Similar to the sequential linear model on students' situational interest, Table 78 summarises the selection and order of expected predictors for the dependent variable 'curiosity towards particle physics'. Here, the variable 'curiosity trait' is entered first because the trait component of curiosity is assumed to be an important predictor of the state component (Cors, 2016; Naylor, 1981).

Table 78. Order of predictors for the sequential regression on curiosity towards particle physics.

#	Variable	Label	Justification
	Novelty Influence Factors	NIFs	
1	Curiosity trait	CT	The trait component of curiosity is an important predictor of the state component (Cors, 2016; Naylor, 1981)
2	Curiosity state particle physics (pre-visit questionnaire)	CS1	The curiosity state at the pre-visit questionnaire is a strong predictor of the post-visit questionnaire results
3	Dispositional physics interest	IE1	Epistemic curiosity is very similar to the construct of interest (Alexander & Grossnickle, 2016)
4	Experience with hands-on experiments	EH	See Table 75
5	English skills (binary)	EN	
6	Experience with OSLEOs	EO	
7	Experience with particle phys.	EP	
	Novelty Experience Factors	NEFs	
8	Tinkering support (novelty)	TIN	
9	Setting orientation (novelty)	SO	
10	Lack of problems (novelty)	PRO	
11	Cognitive load experiments	CLE	
12	Tutor support	TS	

Evaluation of assumptions of sequential linear regression

The evaluation of the assumptions of linear modelling is summarised in Table 79. Moreover, the residual scatterplot is included in Appendix 10.9.3. As before, the normality and homoscedasticity assumptions of the regression residuals were violated. Therefore, the standard errors of the predictors might be underestimated, and a smaller alpha level may be appropriate when evaluating significant predictors by gauging their p-values. However, the Bonferroni correction with $m = 12$ was already applied when evaluating the 12 predictors, and this conservative correction already reduces the critical alpha level significantly. Also, 23 outliers were removed from the sample.

Table 79. Assumptions of sequential linear regression and results of their evaluation.

Assumptions	Target (see 5.5.2)	Main study result	
Ratio of cases to independent variables	$N \geq 50 + 8m$ $N \geq 104 + m$	$m = 12, N = 509$ $N \geq 146$	✓
Absence of multivariate outliers	Mahalanobis distance < 33 $(df = 12, \alpha = 0.001)$	18 critical cases[18]	↯
Normality	Rectangular scatterplot: z(regression residuals) over z(predicted regression values)	Strong ceiling effect for high z-values	↯
Homoscedasticity			↯
Linearity		OK	✓
Independent residuals	Durbin Watson $1.5 < DW < 2.5$	1.9	✓
Outliers in the solution	Standard. residuals $\lvert z \rvert \leq 3.29$	5 critical cases[19]	↯
Multicollinearity	correlation $\lvert r \rvert < 0.70$	< 0.59	✓
Multicollinearity	tolerance > 0.10	> 0.47	✓

Appendix 10.9.3 lists the full correlation matrix for all predictors and the dependent variable. High correlations ($r > 0.5$) were identified between 'curiosity trait' and 'curiosity towards particle physics' measured in the pre-visit questionnaire ($r_{CT,CS1} = 0.52$, $p = 10^{-36}$). This correlation is consistent with the theoretical formulation of curiosity trait as a potential to experience curiosity states (Naylor, 1981). Also, as expected, the curiosity state before taking part in the hands-on session correlates highly with the curiosity state after the hands-on session in S'Cool LAB ($r_{CS1,CS3} = 0.57,\ p = 6 \cdot 10^{-46}$).

Furthermore, as expected, 'dispositional interest' and 'curiosity towards particle physics' were highly correlated, both with the curiosity state before the hands-on sessions in S'Cool LAB ($r_{IE1,CS1} = 0.65,\ p = 10^{-63}$) and only a little less with the curiosity state after the hands-on sessions ($r_{IE1,CS3} = 0.48,\ p = 3 \cdot 10^{-31}$) because they are very similar constructs (Alexander & Grossnickle, 2016). Moreover, a moderate correlation was identified between 'curiosity trait' and 'dispositional physics interest' ($r_{CT,IE1} = 0.43$, $p = 6 \cdot 10^{-24}$). This correlation cannot be explained by the theoretical background of the constructs interest and curiosity. One could hypothesise that students who are interested in physics are essentially curious about the universe, which could correlate with general curiosity towards new things. Another moderate but interesting correlation was identified between 'curiosity towards particle physics' in the pre-visit questionnaire and perceived 'tinkering support' in the post-visit questionnaire ($r_{CS1,TIN} = 0.38,\ p = 3 \cdot 10^{-19}$). Here, students who are curious about the topic particle physics also appreciated the interactions with the tutoring particle physicists more. This correlation appears also in the post-visit

[18] Multivariate outliers: cases 72, 85, 133, 146, 157, 187, 201, 253, 269, 320, 328, 353, 365, 378, 380, 398, 410, 439

[19] Outliers in the solution: cases 198, 287, 386, 434, 478

questionnaire between the respective curiosity state and 'tutor support' ($r_{CS3,TIN} = 0.48$, $p = 3 \cdot 10^{-31}$).

Further significant correlations were identified among all of the novelty experience factors, and between all NEFs and the dependent variable. However, all correlations were below the critical value of 0.70 and tolerance values > 0.5 confirmed that there were no critical multicollinearity issues for the data set. Nevertheless, the order in which these highly correlated predictors entered the linear model might have affected the size of their regression coefficients due to their shared variance. In particular, the regression coefficients for the different novelty experience factors might be difficult to interpret.

Results and discussion

Table 80 displays the standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$). In the following, only the solution without outliers is interpreted. With all 12 independent variables the model accounted for a significant amount of 49% of the variance in students' curiosity towards particle physics ($Adjusted\ R^2 = 0.49, SEE = 0.67, F(12,473) = 40, p = 2 \cdot 10^{-64}$), which indicated a moderate model fit for the data according to Muijs (2004).

The stable disposition 'curiosity trait' alone accounted for 18% of the variance of the dependent variable. Adding students' curiosity state towards particle physics before the hands-on session increased the adjusted R^2 significantly to 39%. However, the model did not improve when adding students' dispositional physics interest and prior experiences (with hands-on experiments, OSLEOs, and particle physics) or their (dichotomised) English skills. Therefore, all novelty influence factors together accounted for 39% of the variance in the dependent variable 'curiosity towards particle physics'. However, even after controlling for novelty influence factors, adding novelty experience factors improved the prediction model. Among the novelty experience factors, tinkering support increased the adjusted R^2 significantly to 46% and students' perceived support by tutors added another 3%.

In total, only three of the 12 regression coefficients differed significantly from zero. As expected, the strongest predictor was students' initial curiosity towards particle physics ($\beta_{w/o} = 0.40^{***}$). Furthermore, the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.16^{*}$), and 'tutor support' ($\beta_{w/o} = 0.22^{***}$) improved the prediction significantly. While curiosity state with respect to particle physics was an important predictor of students' situational interest in S'Cool LAB's particle physics activities, the opposite was not the case: students' dispositional physics interest was no significant predictor of students'

curiosity state with respect to particle physics after the activities ($\beta_{w/o} = 0.08^{ns}$), although there was a relatively high correlation between the two variables. This could reflect multicollinearity issues with the previous predictor 'curiosity towards particle physics' (pre), which accounted for the overlapping variance.

These results suggest that higher levels of curiosity towards particle physics were reached by students who were already curious about the topic before coming to CERN and S'Cool LAB. Furthermore, students who perceived the learning atmosphere favourable with high levels of tinkering and tutor support benefited more from the hands-on session with respect to the educational outcome curiosity state. However, these relationships may be merely correlational: students who were curious about particle physics appreciated the contact to particle physicists more than less curious students.

Table 80. Sequential regression of novelty influence and experience factors on students' curiosity towards particle physics: standardised regression coefficients and p-values for the full sample (β_{all},p_{all}), and for the reduced sample without outliers ($\beta_{w/o}$,$p_{w/o}$), 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[20] and significant changes in the adjusted R^2 ($p < 0.004$) are highlighted.

	full sample		w/o outliers			95% CI		
	$N = 509$		$N = 485$			Lower	Upper	Adj.
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)		0.277			0.127	-0.17	0.02	
Curiosity trait CT	0.04	0.302	0.09		0.014	0.02	0.18	**0.18**
Curiosity state particle phys. CS1	0.33	3E-11	**0.40**	***	5E-17	0.32	0.51	**0.39**
Dispositional physics interest IE1	0.12	0.008	0.08		0.065	0.00	0.17	0.40
Experience hands-on experim. EH	-0.01	0.732	-0.01		0.762	-0.07	0.05	0.40
English skills (binary) EN	0.05	0.167	0.06		0.080	-0.01	0.24	0.40
Experience with OSLEOs EO	0.01	0.792	-0.01		0.729	-0.10	0.07	0.40
Experience with particle physics EP	-0.01	0.727	-0.03		0.327	-0.09	0.03	0.40
Tinkering support (novelty) TIN	0.21	1E-05	**0.16**	*	1E-03	0.06	0.25	**0.46**
Setting orientation (novelty) SO	0.06	0.139	0.02		0.688	-0.06	0.09	0.46
Lack of problems (novelty) PRO	-0.07	0.104	-0.05		0.193	-0.13	0.03	0.46
Cognitive load experiments CLE	-0.02	0.720	0.01		0.903	-0.08	0.09	0.46
Tutor support TS	0.16	1E-04	**0.22**	***	3E-07	0.15	0.33	**0.49**
Full model:	$Adjusted\ R^2 = 0.44$		$Adjusted\ R^2 = 0.49, SEE = 0.67$ $F(12,473) = 40, p = 2 \cdot 10^{-64}$					

[20] The Bonferroni correction reduces the required p-values with $m = 13$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

7.7.3 SLM: Perceived enjoyment and helpfulness

Similar to the sequential linear model on students' situational interest, Table 81 summarises the selection and order of expected predictors for the dependent variables 'enjoyment' and 'helpfulness'. Here, the two regression models are described together, because of their similarities and their link to the interest construct. In particular, enjoyment is sometimes even used interchangeably with interest (Renninger, 2007).

'Helpfulness', on the other hand, summarises the perceived usefulness of the activities for the students (e.g., "*The time in S'Cool LAB helped me to get better at doing experiments*.") In this way, 'helpfulness' was closely related to the value-oriented component of interest and served as an indicator for a hold-component of interest development.

Table 81. Order of predictors for the sequential regressions on enjoyment and helpfulness.

#	Variable	Label	Justification
	Novelty Influence Factors	NIFs	
1	Dispositional physics interest	IE1	**Enjoyment** and interest share similarities (Renninger, 2007), therefore the dispositional interest could be an important predictor. **Helpfulness** shares similarities with the 'hold' component of interest (Krapp, 2002), therefore the dispositional interest could be an important predictor.
2	Experience with hands-on experiments	EH	See Table 75
3	English skills (binary)	EN	
4	Experience with OSLEOs	EO	
5	Experience with particle phys.	EP	
	Novelty Experience Factors	NEFs	
6	Tinkering support (novelty)	TIN	
7	Setting orientation (novelty)	SO	
8	Lack of problems (novelty)	PRO	
9	Cognitive load experiments	CLE	
10	Tutor support	TS	

Evaluation of assumptions of sequential linear regression

The evaluation of the assumptions of linear modelling is summarised in Table 82. Moreover, the residual scatterplot is included in Appendix 10.9.4. As before, the normality and homoscedasticity assumptions of the regression residuals were violated for both regression models. Therefore, the standard errors of the predictors might be underestimated, and a smaller alpha level may be appropriate when evaluating significant predictors by gauging their p-values. However, the Bonferroni correction with $m = 10$ was applied when evaluating the 10 predictors, and this conservative correction already

reduces the critical alpha level significantly. Also, 16 outliers were removed from the sample in each of the two regressions.

Appendix 10.9.4 lists the full correlation matrix for all predictors and the two dependent variables. Significant correlations were identified among all of the novelty experience factors, and between all NEFs and the dependent variables enjoyment and helpfulness with only one exception: the novelty experience component 'lack of problems' did not correlate with the perceived helpfulness of the activities. High correlations ($r > 0.5$) were identified between the novelty experience factor 'tinkering support' and the outcomes 'enjoyment' ($r_{TIN,ENJ} = 0.64,\ p = 7 \cdot 10^{-59}$) as well as 'helpfulness' ($r_{TIN,HEL} = 0.52,\ p = 3 \cdot 10^{-37}$). Also, the novelty experience factor 'tutor support' correlated highly with the outcome 'enjoyment' ($r_{TS,ENJ} = 0.55,\ p = 3 \cdot 10^{-42}$) and moderately with the outcome 'helpfulness' ($r_{TS,HEL} = 0.42,\ p = 3 \cdot 10^{-23}$). In general, all correlations between predictors were below the critical value of 0.70 and tolerance values > 0.5 confirmed that there were no critical multicollinearity issues for the data set. Nevertheless, the order in which these highly correlated predictors entered the linear model might have affected the size of their regression coefficients due to their shared variance. In particular, the regression coefficients for the different novelty experience factors might be difficult to interpret.

Table 82. Assumptions of sequential linear regression and results of their evaluation.

Assumptions	Target (see 5.5.2)	Result enjoyment		Result helpfulness	
Ratio of cases to independent variables	$N \geq 50 + 8m$ $N \geq 104 + m$	$m = 10$ $N = 509$ $N \geq 130$	✓	$m = 10,$ $N = 509$ $N \geq 130$	✓
Absence of multivariate outliers	Mahal. distance < 30 $(df = 10, \alpha = 0.001)$	13 critical cases[21]	↯	13 critical cases[22]	↯
Normality	Scatterplot: z(regression residuals) over z(predicted regression values)	Strong ceiling effect for high z-values	↯	Ceiling effect for high z-values	↯
Homoscedasticity			↯		↯
Linearity		OK	✓	OK	✓
Independent residuals	Durbin Watson $1.5 < DW < 2.5$	2.0	✓	1.8	✓
Outliers in the solution	Standard. residuals $\|z\| \leq 3.29$	3 critical cases[23]	↯	3 critical cases[24]	↯
Multicollinearity	correlation < 0.70	< 0.59	✓	< 0.59	✓
Multicollinearity	tolerance > 0.1	> 0.51	✓	> 0.51	✓

[21] Multivariate outliers: cases 72, 85, 133, 146, 187, 253, 320, 328, 365, 378, 394, 410, 439
[22] Multivariate outliers: cases 72, 85, 133, 146, 187, 253, 320, 328, 365, 378, 394, 410, 439
[23] Outliers in the solution: cases 161, 266, 363
[24] Outliers in the solution: cases 167, 328, 448

Two moderate but interesting correlations were identified between students' 'dispositional physics interest' in the pre-visit questionnaire and the outcomes 'enjoyment' ($r_{IE1,ENJ} = 0.34,\ p = 1 \cdot 10^{-15}$) and 'helpfulness' ($r_{IE1,HEL} = 0.33,\ p = 2 \cdot 10^{-14}$). Here, students who are interested in physics in general enjoyed the activities in S'Cool LAB more and also perceived them as more useful than students with lower physics interest. These correlations were expected because of the theoretical background of the interest construct and its relationship to the variables 'enjoyment' and 'helpfulness'.

Results and discussion for regression on enjoyment

Table 83 displays the standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the sample without outliers ($\beta_{w/o}, p_{w/o}$). In the following, only the solution without outliers is interpreted. With all 10 independent variables the model accounted for a significant amount of 48% of the variance in students' enjoyment ($Adjusted\ R^2 = 0.48, SEE = 0.67, F(10,482) = 46, p = 2 \cdot 10^{-64}$), which indicates a good model fit for the data according to Muijs (2004).

Table 83. Sequential regression of novelty influence and experience factors on students' enjoyment: standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[25] and significant changes in the adjusted R^2 ($p < 0.004$) are highlighted.

	full sample		w/o outliers			95% C. I.		
	$N = 509$		$N = 492$			Lower	Upper	Adj.
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)		0.263			0.298	-0.05	0.15	
Dispositional physics interest IE1	0.14	2E-04	**0.12**	**	9E-04	0.05	0.19	**0.10**
Experience hands-on experim. EH	-0.03	0.437	-0.04		0.226	-0.10	0.02	0.10
English skills (binary) EN	-0.05	0.155	-0.04		0.259	-0.20	0.05	0.10
Experience with OSLEOs EO	0.00	0.920	-0.02		0.465	-0.12	0.05	0.10
Experience with particle physics EP	-0.02	0.489	-0.04		0.262	-0.10	0.03	0.10
Tinkering support (novelty) TIN	0.41	1E-18	**0.37**	***	2E-14	0.26	0.44	**0.41**
Setting orientation (novelty) SO	0.05	0.227	0.05		0.219	-0.03	0.12	0.42
Lack of problems (novelty) PRO	-0.02	0.626	-0.04		0.306	-0.12	0.04	0.41
Cognitive load experiments CLE	-0.06	0.177	-0.05		0.267	-0.13	0.04	0.42
Tutor support TS	0.27	2E-11	**0.33**	***	4E-14	0.26	0.44	**0.48**
Full model:	$R^2 = 0.48$		$Adjusted\ R^2 = 0.48, SEE = 0.67$ $F(10,482) = 46, p = 2 \cdot 10^{-64}$					

The predictor 'dispositional physics interest' accounted for 10% of the variance of students' enjoyment. However, the model did not improve when adding students' prior experiences

[25] The Bonferroni correction reduces the required p-values with $m = 10$ to
* $p < 0.005$, ** $p < 0.001$ and *** $p < 0.0001$

(with hands-on experiments, OSLEOs, and particle physics) and their (dichotomised) English skills. Therefore, all novelty influence factors together accounted for only 10% of the variance in the dependent variable 'enjoyment'. After controlling for novelty influence factors, adding novelty experience factors greatly improved the prediction model. Among the novelty experience factors, 'tinkering support' increased the adjusted R^2 significantly to 41% and students' perceived support by tutors added another 7%.

In total, only three of the 10 regression coefficients differed significantly from zero. The two strongest predictors were the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.37^{***}$), and 'tutor support' ($\beta_{w/o} = 0.33^{***}$). Furthermore, students' dispositional interest was the third significant but weaker predictor ($\beta_{w/o} = 0.12^{**}$).

Results and discussion of regression on helpfulness

Similar results were found for students' perceived helpfulness of the S'Cool LAB activities. Here, Table 84 displays the standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$). As for the previous regression models, only the solution without outliers was interpreted. With all 10 independent variables the model accounted for a significant amount of 39% of the variance in the dependent variable 'perceived helpfulness* ($Adjusted\ R^2 = 0.39, SEE = 0.75, F(10{,}483) = 33, p = 7 \cdot 10^{-49}$), which indicates a moderate model fit for the data according to Muijs (2004).

The predictor 'dispositional physics interest' accounted for 11% of the variance of the dependent variable. As for the outcome 'enjoyment', the model did not improve when adding students' prior experiences or English skills. Therefore, all novelty influence factors together accounted for only 11% of the variance. After controlling for novelty influence factors, adding novelty experience factors greatly improved the prediction model. Among the novelty experience factors, 'tinkering support' increased the adjusted R^2 significantly to 33%. Including the 'lack of problems' as predictor increased R^2 to 35% and students' perceived support by tutors added another 4%.

In total, four of the 10 regression coefficients differed significantly from zero. Here, students' dispositional interest was a significant but relatively weak predictor of students' perceived helpfulness ($\beta_{w/o} = 0.16^{***}$). The two strongest predictors were the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.36^{***}$), and 'tutor support' ($\beta_{w/o} = 0.27^{***}$).

Table 84. Sequential regression of novelty influence and experience factors on students' perceived helpfulness: standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[26] and significant changes in the adjusted R^2 ($p < 0.004$) are highlighted.

	full sample		w/o outliers			95% C. I.		
	$N = 509$		$N = 493$			Lower	Upper	Adj.
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)		5E-14			6E-10	-3.85	-2.03	
Dispositional physics interest IE1	0.16	8E-05	**0.16**	***	5E-05	0.08	0.24	**0.11**
Experience hands-on experim. EH	0.05	0.131	0.07		0.066	0.00	0.13	0.11
English skills (binary) EN	-0.04	0.292	-0.02		0.563	-0.18	0.10	0.11
Experience with OSLEOs EO	0.00	0.985	0.00		0.946	-0.10	0.09	0.11
Experience with particle physics EP	-0.10	0.005	-0.07		0.047	-0.14	0.00	0.11
Tinkering support (novelty) TIN	0.38	9E-14	**0.36**	***	8E-12	0.42	0.74	**0.33**
Setting orientation (novelty) SO	0.13	0.002	0.09		0.041	0.00	0.22	0.34
Lack of problems (novelty) PRO	-0.15	0.001	**-0.16**	**	5E-04	-0.22	-0.06	**0.35**
Cognitive load experiments CLE	0.02	0.641	0.01		0.781	-0.08	0.11	0.35
Tutor support TS	0.19	2E-05	**0.27**	***	8E-09	0.20	0.39	**0.39**
Full model:	$R^2 = 0.36$		$Adjusted\ R^2 = 0.39, SEE = 0.75$ $F(10{,}483) = 33, p = 7 \cdot 10^{-49}$					

As a third novelty experience factor, lack of problems ($\beta_{w/o} = -0.16^{**}$) was a significant predictor. Although the size of the standardised regression coefficient is rather small, its minus sign and p-value are surprising. While the three items of this scale were negatively phrased (e.g., *"I was afraid to make mistakes…"*), they were coded inversely to maintain the comparability with the other novelty experience items. Therefore, the minus sign indicates that students perceived the activities as more helpful, if they struggled more with the activities. However, there was no significant correlation between the two variables 'lack of problems' and 'helpfulness' ($r_{PRO,HEL} = 0.04,\ p = 0.315$).

This supports the assumption of one (or more) suppressor variables being present in the regression model. Indeed, a suppressor variable can increase the regression coefficient of another variable when it is included in a regression model (MacKinnon, Krull, & Lockwood, 2000). Accordingly, only when 'lack of problems' enters the model together with a third variable, its regression coefficient becomes different from zero because the suppressor variable suppresses variance in the predictor (here: 'lack of problems'). In this case, the direct interpretation of the regression coefficient becomes difficult.

[26] The Bonferroni correction reduces the required p-values with $m = 10$ to
* $p < 0.005$, ** $p < 0.001$ and *** $p < 0.0001$

To find out, which of the novelty experience factors acted as suppressor variable in this case, a series of regression models were calculated with the dependent variable 'helpfulness' and the predictor 'lack of problems' as fixed variables while adding the other novelty experience variables one at the time as third variable. In these models, the changes in the regression coefficients were analysed as suggested by MacKinnon et al. (2000). Here, 'tinkering support' was identified as strongest suppressor variable, while 'cognitive load experiments' and 'tutor support' showed a tendency of suppression effects, see Table 85. Consequently, the predictor 'lack of problems' should not be interpreted in the regression on perceived helpfulness.

These results suggest that higher levels of perceived enjoyment and helpfulness were reached by students who were generally interested in physics. This supports the hypothetical relationships between the variables with 'enjoyment' being an indicator of interest and 'helpfulness' serving as indicator of the 'hold' component of interest. However, more importantly, students who perceived the learning atmosphere favourable with high levels of tinkering and tutor support also perceived the activities in S'Cool LAB as more enjoyable and helpful.

Table 85. Series of linear regression models on students' perceived helpfulness with the predictor 'lack of problems'. Potential confounding variables were entered separately to determine their effect on the regression coefficient and p-value of the predictor 'lack of problems'.

		$N = 493$			suppression effect
Model	**Predictors**	β_{adj}		p_{adj}	$\beta_{adj_0} - \beta_{adj_x}$
0	Lack of problems (novelty) PRO	0.05		0.295	
1	Lack of problems (novelty) PRO	**-0.15**	**	2E-04	0.20
	Tinkering support (novelty) TIN	**0.57**	***	4E-38	
2	Lack of problems (novelty) PRO	-0.02		0.604	0.07
	Setting orientation (novelty) SO	**0.37**	***	2E-16	
3	Lack of problems (novelty) PRO	-0.11		0.052	0.16
	Cognitive load experiments CLE	**-0.26**	***	5E-06	
4	Lack of problems (novelty) PRO	-0.10		0.015	0.15
	Tutor support TS	**0.50**	***	2E-29	

7.7.4 SLM: Situational self-beliefs

Table 86 summarises the selection and order of expected predictors for a sequential linear regression on students' situational self-beliefs. The control variables gender and age were entered first into the model because they may affect the succeeding predictors. In particular, girls often report lower values for physics self-beliefs (Hoffmann et al., 1997).

Table 86. Order of predictors for the sequential regression on situational self-beliefs.

#	Variable	Label	Justification
1	Gender (femaleness)	SEX	Girls tend to report lower values for physics self-beliefs (Hoffmann et al., 1997)
2	Age	AGE	Self-beliefs are formed through experience
	Novelty Influence Factors	NIFs	
3	Dispositional self-beliefs	SC1	Dispositional self-beliefs are the most important predictors of situational components as self-beliefs are formed through experience (Gutman & Schoon, 2013)
4	Experience with hands-on experiments	EH	See Table 75
5	English skills (binary)	EN	
6	Experience with OSLEOs	EO	
7	Experience with particle phys.	EP	
	Novelty Experience Factors	NEFs	
8	Tinkering support (novelty)	TIN	
9	Setting orientation (novelty)	SO	
10	Lack of problems (novelty)	PRO	
11	Cognitive load experiments	CLE	
12	Tutor support	TS	

Evaluation of assumptions of sequential linear regression

The evaluation of the assumptions of linear modelling is summarised in Table 87. Moreover, the residual scatterplot is included in Appendix 10.9.5.

Table 87. Assumptions of sequential linear regression on self-beliefs and results of their evaluation.

Assumptions	Target (see 5.5.2)	Main study result	
Ratio of cases to independent variables	$N \geq 50 + 8m$ $N \geq 104 + m$	$m = 12, N = 509$ $N \geq 146$	✓
Absence of multivariate outliers	Mahalanobis distance < 33 $(df = 12, \alpha = 0.001)$	11 critical cases[27]	↯
Normality	Scatterplot: z(regression residuals) over z(predicted regression values)	Ceiling effect for high z-values	↯
Homoscedasticity			↯
Linearity		OK	✓
Independent residuals	Durbin Watson $1.5 < DW < 2.5$	1.8	✓
Outliers in the solution	Standard. residuals $\lvert z \rvert \leq 3.29$	5 critical cases[28]	↯
Multicollinearity	correlation $\lvert r \rvert < 0.70$	< 0.59	✓
Multicollinearity	tolerance > 0.10	> 0.53	✓

As before, the normality and homoscedasticity assumptions of the regression residuals were violated. Therefore, the standard errors of the predictors might be underestimated

[27] Multivariate outliers: cases 72, 85, 146, 157, 187, 253, 320, 328, 365, 410, 439

[28] Outliers in the solution: cases 29, 146, 187, 378, 434

and a smaller alpha level may be appropriate when evaluating significant predictors by gauging their p-values. However, the Bonferroni correction with $m = 12$ was applied when evaluating the 12 predictors, and this conservative correction already reduces the critical alpha level significantly. Also, 14 outliers were removed from the sample.

Appendix 10.9.5 lists the full correlation matrix for all predictors and the dependent variable. Moderate to high correlations were identified between all novelty experience factors and the dependent variable 'situational self-beliefs, for example, the perceived cognitive load ($r_{CLE,SC3} = -0.50,\ p = 2 \cdot 10^{-33}$) or the perceived tutor support ($r_{TS,SC3} = 0.45,\ p = 3 \cdot 10^{-27}$). Furthermore, a moderate correlation was identified between students' 'dispositional physics self-beliefs' and their 'situational self-beliefs' ($r_{SC1,SC3} = 0.38,\ p = 3 \cdot 10^{-19}$). As expected, the relatively stable disposition affected the actualised situational component of self-beliefs. A small but interesting correlation was identified between students' dichotomised English skills and their situational self-beliefs ($r_{EN,SC3} = 0.18,\ p = 3 \cdot 10^{-5}$). Indeed, students with better English skills were expected to handle the cognitive challenges provided by S'Cool LAB's workshops better than students who would need to devote a higher proportion of their working memory to process the foreign language. Also, as expected, girls reported slightly lower 'dispositional self-beliefs' ($r_{SEX,SC1} = -0.18,\ p = 4 \cdot 10^{-5}$), while there was no correlation between gender and the situational self-beliefs. Finally, the control variable age did not correlate with students' self-beliefs. Overall, all correlations were below the critical value of 0.70 and tolerance values > 0.5 confirmed that there were no critical multicollinearity issues for the data set. Nevertheless, the order in which these highly correlated predictors enter the linear model might have affected the size of their regression coefficients due to their shared variance.

Results and discussion

Table 88 displays the standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the sample without outliers ($\beta_{w/o}, p_{w/o}$). As in the previous models, removing 14 outliers had a small effect on the size of the regression coefficients and reduced, overall, their standard errors as reflected in smaller p-values. Therefore, only the solution without outliers is interpreted below. With all 12 independent variables the model accounted for 52% of the variance in students' situational self-beliefs ($Adjusted\ R^2 = 0.52, SEE = 0.65, F(12{,}482) = 46, p = 6 \cdot 10^{-72}$), which indicates a good model fit for the data according to Muijs (2004).

While the control variables age and gender did not account for any of the variance in the dependent variable, adding 'dispositional physics self-beliefs' into the model accounted for

16% of the variance of the situational self-beliefs. Adding students' prior experiences to the model did not improve R^2, nor did students' (dichotomized) English skills. Therefore, all novelty influence factors together accounted for only 18% of the variance in the dependent variable 'situational self-beliefs'. However, even after controlling for novelty influence factors, adding novelty experience factors improved the prediction. Indeed, every single novelty experience factor accounted for a significant amount of additional variance. For example, adding 'tinkering support' increased R^2 significantly to 37%. Adding the other NEFs accounted for additional 15% of the variance in the dependent variable.

Table 88. Sequential regression of novelty influence and experience factors on students' situational self-beliefs: standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[29] and significant changes in the adjusted R^2 ($p < 0.004$) are highlighted.

	full sample		w/o outliers			95% CI		
	$N = 509$		$N = 495$			Lower	Upper	Adj.
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)		0.144			0.021	-1.65	-0.13	
Gender	-0.04	0.242	0.00		0.963	-0.13	0.12	0.00
Age	0.02	0.472	0.02		0.466	-0.04	0.08	0.00
Dispositional self-beliefs SC1	0.19	2E-07	**0.21**	***	2E-09	0.14	0.27	**0.16**
Experience hands-on experim. EH	0.01	0.702	0.00		0.958	-0.06	0.06	0.16
English skills (binary) EN	0.09	0.006	0.06		0.055	0.00	0.24	0.17
Experience with OSLEOs EO	0.05	0.126	0.03		0.285	-0.04	0.13	0.17
Experience with particle phys. EP	-0.03	0.408	-0.02		0.503	-0.08	0.04	0.18
Tinkering support (novelty) TIN	0.06	0.176	0.10		0.036	0.01	0.30	**0.37**
Setting orientation (novelty) SO	0.23	2E-08	**0.25**	***	4E-10	0.16	0.31	**0.43**
Lack of problems (novelty) PRO	0.04	0.283	0.02		0.570	-0.05	0.10	**0.45**
Cognitive load experiments CLE	-0.24	3E-08	**-0.31**	***	7E-13	-0.38	-0.22	**0.51**
Tutor support TS	0.24	7E-09	**0.16**	***	1E-04	0.08	0.25	**0.52**
Full model:	$R^2 = 0.46$		$Adjusted\ R^2 = 0.52, SEE = 0.65$ $F(12{,}483) = 46, p = 6 \cdot 10^{-72}$					

In total, only four of the 12 regression coefficients differed significantly from zero. The strongest predictor of situational self-beliefs among the novelty influence factors were students' dispositional self-beliefs ($\beta_{w/o} = 0.21^{***}$). Among the novelty experience factors, the strongest predictors were 'cognitive load experiments' ($\beta_{w/o} = -0.31^{***}$), 'setting orientation due to preparation' ($\beta_{w/o} = 0.25^{***}$), and 'tutor support' ($\beta_{w/o} = 0.16^{***}$). The

[29] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

predictors 'lack of problems' and 'tinkering support' did not appear as significant predictors. Here, the other novelty experience factors might account for the shared variance.

These results suggest that higher situational self-beliefs with respect to the S'Cool LAB activities were reached by students who had high dispositional physics self-beliefs. Furthermore, students who perceived the learning atmosphere in S'Cool LAB as favourable with high levels of setting orientation and tutor support (and low levels of cognitive load related to the experiments) benefited more from the hands-on workshops with respect to the educational outcome situational self-beliefs. In particular, novelty experience factors contributed significantly and accounted for most of the variance in students' situational self-beliefs even after controlling for novelty influence factors. This dominance of novelty experience factors might also explain the disappearance of gender effects: girls and boys reported significantly different levels of their dispositional self-beliefs in the pre-visit questionnaire, but not in their situational self-beliefs the post-visit questionnaire. However, as the dispositional self-beliefs accounted for only 16% of the variance of the situational self-beliefs, this small gender difference disappeared in this regression model. Furthermore, no significant interaction effects between gender and the predictor variables were found in the outlier-corrected sample.

7.7.5 SLM: Perceived easiness

Table 89 summarises the selection and order of expected predictors for a sequential linear regression on students' perceived easiness of the S'Cool LAB activities. Here, perceived easiness was expected to correlate strongly with students' self-beliefs. Therefore, students' dispositional physics self-beliefs were included first in the regression model.

Table 89. Order of predictors for the sequential regression on perceived easiness.

#	Variable	Label	Justification
	Novelty Influence Factors	NIFs	
1	Dispositional self-beliefs	SC1	As 'easiness' is expected to strongly correlate with self-beliefs, disp. self-beliefs could be an important predictor
2	Experience with hands-on experiments	EH	See Table 75
3	English skills (binary)	EN	
4	Experience with OSLEOs	EO	
5	Experience with particle phys.	EP	
	Novelty Experience Factors	NEFs	
6	Tinkering support (novelty)	TIN	
7	Setting orientation (novelty)	SO	
8	Lack of problems (novelty)	PRO	
9	Cognitive load experiments	CLE	
10	Tutor support	TS	

Evaluation of assumptions of sequential linear regression

The evaluation of the assumptions of linear modelling is summarised in Table 90. Moreover, the residual scatterplot is included in Appendix 10.9.6. As before, the normality and homoscedasticity assumptions of the regression residuals were violated. However, the Bonferroni correction with $m = 10$ was applied when evaluating the 10 predictors, and this conservative correction already reduces the critical alpha level significantly. Also, 15 outliers were removed from the sample.

Table 90. Assumptions of sequential linear regression on easiness and results of their evaluation.

Assumptions	Target (see 5.5.2)	Main study result	
Ratio of cases to independent variables	$N \geq 50 + 8m$ $N \geq 104 + m$	$m = 10, N = 509$ $N \geq 130$	✓
Absence of multivariate outliers	Mahalanobis distance < 35 $(df = 13, \alpha = 0.001)$	14 critical cases[30]	↯
Normality	Scatterplot: z(regression residuals) over z(predicted regression values)	Ceiling effect for high z-values	↯
Homoscedasticity			↯
Linearity		OK	✓
Independent residuals	Durbin Watson $1.5 < DW < 2.5$	1.9	✓
Outliers in the solution	Standard. residuals$\lvert z \rvert \leq 3.29$	1 critical case[31]	↯
Multicollinearity	correlation < 0.70	< 0.59	✓
Multicollinearity	tolerance > 0.1	> 0.5	✓

Appendix 10.9.6 lists the full correlation matrix for all predictors and the dependent variable. Moderate to high correlations were identified between all novelty experience factors and the dependent variable, for example, the perceived cognitive load ($r_{CLE,EAS} = -0.61,\ p = 1 \cdot 10^{-53}$) or the perceived tinkering support ($r_{TIN,EAS} = 0.55,\ p = 2 \cdot 10^{-41}$). A small but interesting correlation was identified between students' experience with particle physics and the perceived easiness ($r_{EP,EAS} = 0.17,\ p = 2 \cdot 10^{-4}$). Indeed, students with more knowledge about and experience in particle physics were expected to perceive S'Cool LAB's workshops as easier. Also, as expected, students' 'dispositional self-beliefs' correlated with the perceived easiness of the S'Cool LAB workshops ($r_{SC1,EAS} = 0.17,\ p = 1 \cdot 10^{-4}$). Further significant correlations existed among all of the novelty experience factors, and between all NEFs and the dependent variable. However, all correlations were below the critical value of 0.70 and tolerance values > 0.5 confirmed that there are no critical multicollinearity issues for the data set. Nevertheless, the order in which these highly correlated predictors entered the linear model might have affected the size of their regression coefficients due to their shared variance. In particular,

[30] Multivariate outliers: cases 72, 85, 146, 157, 187, 253, 320, 328, 365, 378, 394, 410, 439, 456

[31] Outliers in the solution: case 284

the regression coefficients for the different novelty experience factors might be difficult to interpret.

Results and discussion

Table 91 displays the standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers $(\beta_{w/o}, p_{w/o})$. As before, removing 15 outliers had a small effect on the size of the regression coefficients and reduced, overall, their standard errors. Therefore, only the solution without outliers is reported below. With all 10 independent variables the model accounted for 52% of the variance in students' perceived easiness $(Adjusted\ R^2 = 0.52, SEE = 0.68, F(10{,}483) = 55, p = 9 \cdot 10^{-74})$, which indicates a good model fit for the data according to Muijs (2004).

Table 91. Sequential regression of novelty influence and experience factors on students' perceived easiness: standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers $(\beta_{w/o}, p_{w/o})$, 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[32] and significant changes in the adjusted R^2 ($p < 0.004$) are highlighted.

	full sample $N = 509$		**w/o outliers** $N = 493$			**95% CI** Lower bound	Upper bound	**Adj.**
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$			**R^2**
(Constant)		1E-18			5E-19	-4.73	-3.08	
Dispositional self-beliefs SC1	-0.04	0.253	-0.05		0.171	-0.11	0.02	**0.02**
Experience hands-on experim. EH	0.02	0.485	0.01		0.760	-0.05	0.07	0.03
English skills (binary) EN	-0.01	0.831	0.00		0.966	-0.13	0.12	0.03
Experience with OSLEOs EO	0.04	0.165	0.04		0.196	-0.03	0.15	0.03
Experience with particle physics EP	0.05	0.146	0.04		0.170	-0.02	0.11	**0.05**
Tinkering support (novelty) TIN	0.23	2E-07	**0.25**	***	4E-08	0.28	0.58	**0.34**
Setting orientation (novelty) SO	0.17	4E-06	**0.17**	***	6E-06	0.13	0.33	**0.37**
Lack of problems (novelty) PRO	0.13	0.001	**0.14**	**	4E-04	0.06	0.20	**0.45**
Cognitive load experiments CLE	-0.36	3E-17	**-0.36**	***	2E-16	-0.46	-0.29	**0.53**
Tutor support TS	0.08	0.040	0.02		0.601	-0.07	0.11	0.52
Full model:	$R^2 = 0.51$		$Adjusted\ R^2 = 0.52, SEE = 0.68$ $F(10{,}483) = 55, p = 9 \cdot 10^{-74}$					

Students' dispositional physics self-beliefs accounted for only 2% of the variance of students' perceived easiness. Among the other novelty influence factors, adding students' prior particle physics experiences significantly improved the model but all influence factors together accounted for only 5% of the variance of the dependent variable.

[32] The Bonferroni correction reduces the required p-values with $m = 10$ to
* $p < 0.005$, ** $p < 0.001$ and *** $p < 0.0001$

However, adding novelty experience factors improved the model. For example, adding 'tinkering support' increased R^2 significantly to 34%. In total, after adding the other NEFs, the model accounted for 52% of the variance of the dependent variable.

In total, only four of the 10 regression coefficients differed significantly from zero, all of them novelty experience factors. The strongest predictors of the perceived easiness were the perceived 'tinkering support' ($\beta_{w/o} = 0.25^{***}$) and the perceived 'cognitive load of the experiments' ($\beta_{w/o} = -0.36^{***}$). Also, 'setting orientation due to preparation' ($\beta_{w/o} = 0.17^{***}$), and 'lack of problems' ($\beta_{w/o} = 0.14^{**}$) had standardised regression coefficients different from zero.

Although students' perception of the support by tutors correlated significantly with the perceived easiness, the regression coefficient for 'tutor support' is negligible. Since 'tutor support' was entered last, the other novelty experience factors might have already accounted for the shared variance. Similarly, students' dispositional self-beliefs and their particle physics experience were no significant predictors although they correlated with the dependent variable. This suggests that some of the novelty experience factors mediated the effects of these two novelty influence factors. When looking at the effect of students' prior particle physics experience, only students' perceived setting orientation was correlated to both this variable and the dependent variable. Indeed, it is plausible that the perceived setting orientation mediates the effect of prior particle physics experience on perceived easiness. In summary, these results suggest that higher levels of perceived easiness are reached by students who perceived the out-of-school learning atmosphere as favourable with high levels of tinkering support, setting orientation, a lack of problems and a low level of cognitive load.

7.7.6 SLM: Perceived cognitive activation

Table 92 summarises the selection and order of expected predictors for a sequential linear regression on students' perceived cognitive activation in S'Cool LAB. Cognitive activation, per definition, is a rather cognitive construct. For example, cognitively activating tasks allow students to connect new information to their prior knowledge (Lipowsky, 2009). But it also closely related to affective variables, such as interest and self-beliefs. Indeed, cognitive activation is described as characteristic of intrinsically motivated actions (Schiefele, 1991). Therefore, cognitive and affective novelty influence factors were considered in the regression model.

Table 92. Order of predictors for the sequential regression on perceived cognitive activation.

#	Variable	Label	Justification
	Novelty Influence Factors	NIFs	
1	Pre-visit score understanding	UN1	NIF (dimension 'cognitive): cognitively activating tasks allow students to connect to their prior knowledge (Lipowsky, 2009)
2	Physics grade	PH	
3	Dispositional physics interest	IE1	Cognitive activation is described as characteristic of intrinsically motivated actions (Schiefele, 1991)
4	Dispositional self-beliefs	SC1	
5	Experience with hands-on experiments	EH	See Table 75
6	English skills (binary)	EN	
7	Experience with OSLEOs	EO	
8	Experience with particle phys.	EP	
	Novelty Experience Factors	NEFs	
9	Tinkering support (novelty)	TIN	
10	Setting orientation (novelty)	SO	
11	Lack of problems (novelty)	PRO	
12	Cognitive load experiments	CLE	
13	Tutor support	TS	

Evaluation of assumptions of sequential linear regression

The evaluation of the assumptions of linear modelling is summarised in Table 93. Moreover, the residual scatterplot is included in Appendix 10.9.7. As before, the normality and homoscedasticity assumptions of the regression residuals were violated. However, the Bonferroni correction with $m = 13$ was applied when evaluating the 13 predictors, and this conservative correction already reduces the critical alpha level significantly. Also, 11 outliers were removed from the sample.

Table 93. Assumptions of sequential linear regression and results of their evaluation.

Assumptions	Target (see 5.5.2)	Main study result	
Ratio of cases to independent variables	$N \geq 50 + 8m$ $N \geq 104 + m$	$m = 13, N = 509$ $N \geq 154$	✓
Absence of multivariate outliers	Mahalanobis distance < 31 $(df = 11, \alpha = 0.001)$	9 critical cases[33]	↯
Normality	Scatterplot: z(regression residuals) over z(predicted regression values)	Not for high and low z-values	↯
Homoscedasticity			↯
Linearity		OK	✓
Independent residuals	Durbin Watson $1.5 < DW < 2.5$	1.8	✓
Outliers in the solution	Standard. residuals $\lvert z \rvert \leq 3.29$	2 critical cases[34]	↯
Multicollinearity	correlation < 0.70	< 0.59	✓
Multicollinearity	tolerance > 0.1	> 0.44	✓

[33] Multivariate outliers: cases 72, 85, 157, 253, 320, 378, 385, 394, 410

[34] Outliers in the solution: cases 130, 383

Appendix 10.9.7 lists the full correlation matrix for all predictors and the dependent variable. Moderate to high correlations were identified between all novelty experience factors and the dependent variable, for example, the perceived 'cognitive load of the experiments' ($r_{CLE,CA} = -0.36,\ p = 1 \cdot 10^{-16}$), 'tinkering support' ($r_{TIN,CA} = 0.60,\ p = 4 \cdot 10^{-50}$) and 'setting orientation' ($r_{SO,CA} = 0.51,\ p = 1 \cdot 10^{-35}$).

A small but interesting correlation was identified between students' latest physics school grade and the perceived cognitive activation ($r_{PH,CA} = 0.17,\ p = 9 \cdot 10^{-5}$). Hence, students with better physics grades perceived a slightly higher cognitive activation. However, students' English skills, prior particle physics experience or the pre-visit score for their conceptual understanding were not correlated with the perceived cognitive activation. Instead, students' dispositional physics self-beliefs correlated with the dependent variable ($r_{SC1,CA} = 0.31,\ p = 1 \cdot 10^{-12}$) as did students' dispositional physics interest ($r_{IE1,CA} = 0.40,\ p = 2 \cdot 10^{-20}$), which supports the theoretical relationship between cognitive activation and intrinsic motivation.

In summary, all correlations were below the critical value of 0.70 and tolerance values > 0.5 confirmed that there were no critical multicollinearity issues for the data set. Nevertheless, the order in which these highly correlated predictors entered the linear model might have affected the size of their regression coefficients due to their shared variance. In particular, the regression coefficients for the different novelty experience factors might be difficult to interpret.

Results and discussion

Table 94 displays the standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{\underline{w}}, p_{\underline{w}}$). Due to the partly missing data in students' conceptual understanding before taking part in the hands-on sessions in S'Cool LAB, the sample size is somewhat smaller than for the other regression models ($N = 392$). Nevertheless, in the following, only the solution without outliers is interpreted, as the differences are not large. With all 13 independent variables the model accounted for 53% of the variance in students' situational self-beliefs ($Adjusted\ R^2 = 0.53$, $SEE = 0.66, F(13{,}379) = 35, p = 2 \cdot 10^{-56}$), which indicates a good model fit for the data according to Muijs (2004).

Among the novelty influence factors, only students' dispositional physics interest accounted for a significant amount of variance. Together with the previously entered variables (conceptual understanding pre UN1 and physics grade PH), this step of the

model accounted for 13% of the variance of cognitive activation. However, adding students' dispositional self-beliefs, their prior experiences and their dichotomized English skills did not improve R^2 significantly. Therefore, all novelty influence factors together accounted for only 14% of the variance in the dependent variable.

Table 94. Sequential regression of novelty influence and experience factors on cognitive activation: standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers $(\beta_{w/o}, p_{w/o})$, 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[35] and significant changes in the adjusted R^2 ($p < 0.004$) are highlighted.

	full sample $N = 403$		w/o outliers $N = 392$			95% CI		
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$	Lower bound	Upper bound	**Adj. R^2**
(Constant)		0.257			0.208	-0.19	0.04	
conceptual understanding pre UN1	-0.06	0.146	-0.05		0.208	-0.12	0.03	0.03
Physics grade PH	0.08	0.086	0.08		0.091	-0.01	0.16	0.03
Dispositional physics interest IE1	0.05	0.321	0.03		0.529	-0.07	0.13	**0.13**
Dispositional self-beliefs SC1	0.10	0.051	0.14		0.009	0.03	0.24	0.13
Experience hands-on experim. EH	-0.04	0.319	-0.02		0.607	-0.09	0.05	0.14
English skills (binary) EN	0.04	0.262	0.06		0.145	-0.04	0.26	0.14
Experience with OSLEOs EO	0.05	0.141	0.07		0.056	0.00	0.19	0.14
Experience with particle physics EP	-0.07	0.045	-0.09		0.013	-0.15	-0.02	0.14
Tinkering support (novelty) TIN	0.35	6E-11	**0.32**	***	2E-09	0.22	0.42	**0.45**
Setting orientation (novelty) SO	0.24	9E-08	**0.22**	***	7E-07	0.13	0.31	**0.49**
Lack of problems (novelty) PRO	-0.05	0.259	-0.07		0.120	-0.16	0.02	0.49
Cognitive load experiments CLE	-0.03	0.462	-0.04		0.437	-0.13	0.06	0.49
Tutor support TS	0.21	8E-06	**0.26**	***	3E-08	0.19	0.39	**0.53**
Full model:	$R^2 = 0.51$		$Adjusted\ R^2 = 0.53, SEE = 0.66$ $F(13{,}379) = 35, p = 2 \cdot 10^{-56}$					

However, adding the novelty experience factor 'tinkering support' significantly improved the model to $R^2 = 0.45$. Including 'setting orientation' accounted for another 4% as did adding students' perceived 'tutor support' with another 4%.

In total, only three of the 13 regression coefficients differed significantly from zero. Although several of the novelty influence factors were correlated with the dependent variable 'cognitive activation', none of their regression coefficients were relevant in the final model. Here, it seems that novelty experience factors mediated the effects of correlated novelty influence factors, such as 'physics grade', 'dispositional physics interest' or 'dispositional physics self-beliefs'. Instead, strong predictors of students' perceived cognitive activation

[35] The Bonferroni correction reduces the required p-values with $m = 13$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

were the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.32^{***}$), 'setting orientation' ($\beta_{w/o} = 0.22^{***}$), and 'tutor support' ($\beta_{w/o} = 0.26^{***}$).

In summary, these results suggest that higher levels of perceived 'cognitive activation' were reached by students who perceived the out-of-school learning atmosphere as favourable with high levels of tinkering and tutor support as well as high levels of setting orientation. Similar to the educational outcome 'perceived easiness', the perceived cognitive activation was not directly predicted by any novelty influence factors. Instead, the novelty experience factors seemed to mediate their effects on the dependent variable.

7.7.7 SLM: Conceptual understanding

Table 95 summarises the selection and order of expected predictors for a sequential linear regression on students' conceptual understanding after taking part in the hands-on session in S'Cool LAB.

Here, the control variable age was entered first because older students with more physics experience were expected to perform better at any standardised test. Afterwards, mainly cognitive novelty influence factors were considered in the regression model. Based on a synthesis of meta-analyses, Hattie (2008) reported a large average effect of $d_C = 1.44$ for self-reported grades on achievement and a medium-sized effect of $d_C = 0.67$ for prior achievement on achievement. Therefore, students' latest physics grade and their pre-visit conceptual understanding were entered next into the linear model.

Table 95. Order of predictors for the sequential regression on conceptual understanding.

#	Variable	Label	Justification
1	Age	AGE	Understanding improves with age
	Novelty Influence Factors	NIFs	
2	Conceptual understanding pre	UN1	NIF (dimension 'cognitive): previous achievement is a strong predictor of achievement (Hattie, 2008)
3	Physics grade	PH	
4	Experience with hands-on experiments	EH	See Table 75
5	English skills (binary)	EN	
6	Experience with OSLEOs	EO	
7	Experience with particle phys.	EP	
	Novelty Experience Factors	NEFs	
8	Tinkering support (novelty)	TIN	
9	Setting orientation (novelty)	SO	
10	Lack of problems (novelty)	PRO	
11	Cognitive load experiments	CLE	
12	Tutor support	TS	

Evaluation of assumptions of sequential linear regression

The evaluation of the assumptions of linear modelling is summarised in Table 96. Moreover, the residual scatterplot is included in Appendix 10.9.8. As before, the normality and homoscedasticity assumptions of the regression residuals were violated. However, the Bonferroni correction with $m = 12$ was applied when evaluating the 12 predictors, and this conservative correction already reduces the critical alpha level significantly. Also, 10 outliers were removed from the sample. Due to the partly missing data in the pre-visit score for students' conceptual understanding (UN1), the sample size was smaller compared to the other regression models ($N = 365$).

Table 96. Assumptions of sequential linear regression and results of their evaluation.

Assumptions	Target (see 5.5.2)	Main study result	
Ratio of cases to independent variables	$N \geq 50 + 8m$ $N \geq 104 + m$	$m = 12, N = 509$ $N \geq 146$	✓
Absence of multivariate outliers	Mahalanobis distance < 31 $(df = 11, \alpha = 0.001)$	10 critical cases[36]	↯
Normality	Scatterplot: z(regression residuals) over z(predicted regression values)	Not for high and low z-values	↯
Homoscedasticity			↯
Linearity		OK	✓
Independent residuals	Durbin Watson $1.5 < DW < 2.5$	1.8	✓
Outliers in the solution	Standard. residuals$\lvert z \rvert \leq 3.29$	No critical case	✓
Multicollinearity	correlation < 0.70	< 0.59	✓
Multicollinearity	tolerance > 0.1	> 0.50	✓

Appendix 10.9.8 lists the full correlation matrix for all predictors and the dependent variable 'conceptual understanding'. As expected, a high correlation was identified between the pre-visit score and the post-visit score for students' conceptual understanding ($r_{UN1,UN3} = 0.55,\ p = 8 \cdot 10^{-31}$). Also, students' physics grade correlated significantly, though rather weakly, with students' conceptual understanding both before ($r_{PH,UN1} = 0.25$, $p = 2 \cdot 10^{-7}$) and after taking part in the hands-on session in S'Cool LAB ($r_{PH,UN3} = 0.33$, $p = 4 \cdot 10^{-13}$). Indeed, these correlations are in line with previous findings, for example, by Hattie (2008), who reported large effects of self-reported grades and prior achievement on achievement.

Also, three of the novelty experience factors correlated weakly with the dependent variable, namely perceived 'cognitive load of the experiments' ($r_{CLE,UN3} = -0.25$, $p = 6 \cdot 10^{-8}$), 'tinkering support' ($r_{TIN,UN3} = 0.20,\ p = 2 \cdot 10^{-5}$) and 'lack of problems' ($r_{PRO,UN3} = 0.28,\ p = 1 \cdot 10^{-9}$). However, both the perceived setting orientation as well as

[36] Multivariate outliers: cases 72, 85, 157, 253, 320, 354, 378, 385, 394, 410

the perceived tutor support did not correlate significantly with the dependent variable. Further small but interesting correlations were identified between students' experience with particle physics (EP) and the dependent variable ($r_{EP,UN3} = 0.20,\ p = 1 \cdot 10^{-5}$) and between students' dichotomised English skills (EN) and the dependent variable ($r_{EN,UN3} = 0.23,\ p = 5 \cdot 10^{-7}$). Indeed, students with prior experience in the topic were expected to achieve higher scores in an achievement test. Also, students with better English skills were expected to benefit more from the English activities in S'Cool LAB.

In summary, all correlations were below the critical value of 0.70 and tolerance values above 0.5 confirmed that there were no critical multicollinearity issues for the data set. Nevertheless, the order of the predictors might have affected the size of their regression coefficients due to shared variance.

Results and discussion

Table 97 displays the standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$). Removing 10 outliers had only a small effect on the size of the regression coefficients. Hence, in the following, only the solution without outliers is reported. With all 12 independent variables the model accounted for 39% of the variance in students' conceptual understanding after the hands-on session ($Adjusted\ R^2 = 0.39, SEE = 0.81, F(12{,}353) = 20, p = 3 \cdot 10^{-33}$) indicating a moderate model fit for the data according to Muijs (2004).

The age of the students did not account for any of the variance in students' conceptual understanding. However, students' pre-visit conceptual understanding accounted for 30% of the variance, and adding the self-reported physics grade accounted for another 4% of the variance in the dependent variable. Afterwards, none of the other novelty influence factors significantly improved the model. Indeed, all novelty influence factors together accounted for 36% of the variance in the dependent variable.

Furthermore, adding novelty experience factors did not improve the regression model significantly with one small exception: students' perceived 'lack of problems' (PRO) significantly increased R^2 from 37% to 39%. The final model accounted for 39% of the variance in the dependent variable.

In total, only two of the 12 regression coefficients differed significantly from zero, students' pre-visit conceptual understanding ($\beta_{w/o} = 0.45^{***}$) and their self-reported physics grade ($\beta_{w/o} = 0.18^{**}$). Although several of the other novelty influence and experience factors were weakly correlated with the dependent variable, none of their regression coefficients

were relevant in the final model. Here, most of the 'unsuccessful predictors' were correlated with the significant predictor 'conceptual understanding pre' (UN1). Therefore, it seems that the latter already accounted for the shared variance.

Table 97. Sequential regression of novelty influence and experience factors on students' situational interest: standardised regression coefficients and p-values for the full sample (β_{all}, p_{all}), and for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), 95% confidence interval (CI) for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[37] and significant changes in the adjusted R^2 ($p < 0.004$) are highlighted.

	full sample		w/o outliers			95% C. I.		
	$N = 375$		$N = 365$			Lower	Upper	Adj.
Predictor	β_{all}	p_{all}	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)		0.323			0.305	-0.22	0.07	
Age	0.02	0.561	0.03		0.556	-0.06	0.12	0.00
Conceptual understanding pre UN1	0.44	1E-21	**0.45**	***	4E-21	0.37	0.55	**0.30**
Physics grade PH	0.18	9E-05	**0.18**	**	9E-05	0.10	0.29	**0.34**
Experience hands-on experim. EH	-0.08	0.076	-0.08		0.052	-0.18	0.00	0.35
English skills (binary) EN	0.07	0.116	0.07		0.107	-0.03	0.35	0.36
Experience with OSLEOs EO	0.04	0.365	0.02		0.614	-0.09	0.15	0.36
Experience with particle physics EP	0.06	0.190	0.05		0.266	-0.04	0.13	0.36
Tinkering support (novelty) TIN	0.08	0.144	0.09		0.155	-0.04	0.23	0.37
Setting orientation (novelty) SO	-0.10	0.045	-0.09		0.060	-0.21	0.00	0.37
Lack of problems (novelty) PRO	0.14	0.007	0.12		0.029	0.01	0.23	**0.39**
Cognitive load experiments CLE	-0.06	0.265	-0.05		0.353	-0.18	0.06	0.39
Tutor support TS	0.03	0.630	0.03		0.543	-0.09	0.17	0.39
Full model:	$R^2 = 0.39$		$Adjusted\ R^2 = 0.39, SEE = 0.81$ $F(12{,}353) = 20, p = 3 \cdot 10^{-33}$					

In summary, these results suggest that higher levels of conceptual understanding were reached by students who had already shown better understanding in the pre-visit questionnaire and who had good physics grades in school. After controlling for these two strong predictors, none of the other novelty influence and experience factors contributed significantly to the model. Compared to the other educational outcomes, it seemed that students' perception of the learning atmosphere in S'Cool LAB was not relevant for their conceptual learning.

7.7.8 Overview of eight sequential linear models on educational outcomes

Table 98 provides an overview of the final eight models including all significant predictors. These sequential linear regression models were calculated to estimate the importance of the different novelty influence factors and novelty experience factors with respect to the

[37] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

various educational outcomes. In particular, novelty influence factors were entered before considering novelty experience factors because novelty experience factors seem to mediate the effect of novelty influence factors on educational outcomes (Cors et al., 2017).

Table 98. Summary of linear regressions on eight educational outcomes. Regression coefficients significantly different from zero are included for relevant predictor variables. Variables that were entered into the respective model but were no significant predictors are marked with an 'x'. Predictors that were significantly correlated with the dependent variable are highlighted in grey. The suppressor effect of the variable 'tinkering support' on the variable 'lack of problems' is highlighted in red and not interpreted here, see 7.7.3.

	Sit. interest	Curiosity state	Enjoyment	Helpfulness	Sit. self-beliefs	Easiness	Cognitive activation	Conceptual understanding
	IE3	CS3	ENJ	HEL	SC3	EAS	CA	UN3
Novelty influence factors	More affective outcomes				More cognitive outcomes			
Gender (SEX)	x				x			
Age (AGE)	x				x			X
Disp. physics interest (IE1)	0.21	x	0.12	0.16			x	
Curiosity trait (CT)		x						
Curiosity state particle (CS1)	0.20	0.40						
Disp. physics self-beliefs (SC1)					0.21	x	x	
Understanding pre (UN1)							x	0.45
Physics grade (PH)							x	0.18
Experience experiments (EH)	x	x	x	x	x	x	x	x
English skills (EN)	x	x	x	x	x	x	x	x
Experience out-of-school (EO)	x	x	x	x	x	x	x	x
Experience particle phys. (EP)	x	x	x	x	x	x	x	x
Novelty experience factors								
Tinkering support (TIN)	0.29	0.16	0.37	0.36	x	0.25	0.32	x
Setting orientation (SO)	x	x	x	x	0.25	0.17	0.22	x
Lack of problems (PRO)	x	x	x	-0.16	x	0.14	x	x
Cognitive load experim. (CLE)	x	x	x	x	-0.31	-0.36	x	x
Tutor support (TS)	0.27	0.22	0.33	0.27	0.16	x	0.26	x
Adjusted R^2	**0.55**	**0.49**	**0.48**	**0.39**	**0.52**	**0.52**	**0.53**	**0.39**
Adjusted R^2 only NIFs	0.31	0.40	0.10	0.11	0.18	0.05	0.14	0.36
Adjusted R^2 only NEFs	0.45	0.29	0.47	0.47	0.48	0.48	0.46	0.09
Removed outliers	16	23	16	15	14	15	11	10

Furthermore, the amount of modelled variance is listed separately in Table 98 for the novelty influence factors and the novelty experience factors[38]. This not only allows to compare the predictive power of the two groups of novelty factors but also helps to estimate the mediation effect of the novelty experience factors. However, before

[38] The linear models based on only novelty experience factors are included in appendix 10.9.

interpreting these results (section 7.9), the single-level sequential linear models are compared with multi-level linear models (section 7.8).

7.8 S'Cool LAB's educational outcomes: multi-level models (MLMs)

Multilevel modelling provides an additional way of analysing data that does not come from randomly selected individuals but from students coming in groups from different countries, within different education systems, from different schools, and educated by different teachers. In addition, multilevel modelling can help to estimate differences between the hands-on sessions.

Although elements like student worksheets or the schedule of the S'Cool LAB Day programme were kept constant throughout the study, the volunteers facilitating the workshops vary between groups. Here, students, who came to CERN on different days might have perceived certain laboratory characteristics (measured as novelty experience factors) differently. Also, since students within one group may share similar prior experiences, their novelty influence factors may not be independent of each other. For example, students from a school that supports science field trips might have shared previous out-of-school learning experiences. Also, different teachers might prepare their classes for out-of-school learning experiences in a specific way (or not at all). Here, contrary to single-level linear regression models, multilevel modelling takes these dependencies into account and in doing so avoids the violation of the assumption of independent observations.

Consequently, students are considered to be level-1 units of multi-level models and 'level-1 variance' refers to variance in certain variables at the student level (i.e., differences between individual students). Moreover, different options of level-2 units are available, namely the grouping variables 'teacher', 'group', and 'country'. Here, 'level-2 variance' refers to variance in certain variables that is estimated at the respective level 2 (i.e., differences that are accounted for by different teachers, groups, or countries). For an introduction into multi-level modelling including its assumptions, equations and parameters, see section 5.6.

In the following, subsection 7.8.1 highlights the need for multi-level modelling by comparing the variance components of S'Cool LAB's educational outcomes. Consequently, subsection 7.8.2 justifies and specifies the use of random-incept models. Then, subsection 7.8.3 describes a series of multi-level models on students' situational interest in detail. In particular, a full multi-level model was calculated first to compare it with the corresponding

single-level model described in subsection 7.7.1. This full multi-level model included the same independent variables and an additional random effect. Afterwards, a series of multi-level models is presented based on the stepwise inclusion of significant predictors. For every step, the variance components, the regression coefficients and the model fit parameters AIC & BIC were computed and compared with the full model. This series of models allows to estimate the effect of predictors. In particular, the series allows to determine how much of the respective level-1 or level-2 variance is reduced when inserting a predictor into the model. Moreover, the model series helps to determine the best model. In a similar way, but described in less detail the other educational outcomes of the S'Cool LAB hands-on sessions are modelled accordingly: curiosity state towards particle physics (7.8.4), perceived enjoyment (7.8.5), perceived helpfulness (7.8.6), situational physics self-beliefs (7.8.7), perceived easiness (7.8.8), perceived cognitive activation (7.8.9), and students' conceptual understanding (7.8.10).

7.8.1 Need for multi-level modelling

Although many studies in physics education research analyse clustered data (e.g., by distributing questionnaires to whole classes of students from different schools), many researchers ignore potential nesting in the datasets and use only single-level linear modelling instead (Van Dusen & Nissen, 2019). However, analysing multi-level data in a single-level model can lead to severe statistical challenges, such as inflated type I errors. Table 99 shows the hierarchical data structure of students taking part in the S'Cool LAB main study. In particular, the 509 students had 69 different physics teachers and were distributed over 28 different groups from 13 different countries.

Table 99. Overview of data structure with number of level-2 clusters and average cluster sizes (number of level-1 units (i.e., students) per level-2 unit) including standard deviation (SD).

	Cluster variable		
	teacher	**group**	**country**
Number of level-2 clusters	69	28	13
Average cluster size (SD)	9 (7)	19 (8)	39 (28)

The variance partition coefficient (VPC)[39] provides a measure to estimate the need for multi-level modelling. As a rule of the thumb, Mehmetoglu and Jakobsen (2016) recommend multi-level modelling if the variance partition coefficient exceeds 5%. This means that more than 5% of the total variance of a dependent variable can be accounted for by differences in the average value of the dependent variable between level-2 units (so-called 'random

[39] The variance partition coefficient (VPC) is the ratio between variance attributed to the group level and the total variance of a dependent variable, which is the sum of the group-level (level-2) variance and the individual level (level-1) variance, see equation 13 in subsection 5.6.1.

intercept') when considering two-level models. Moreover, the variance inflation factor (VIF)[40] describes by which factor the sample size would be overestimated when using a single-level model on clustered data.

Table 100 shows the VPC and VIF for multi-level models of the eight educational outcomes of S'Cool LAB, each for the three different clustering options: teachers, groups and countries. Here, only two-level models were considered to keep the models as simple as possible. Furthermore, the model fit criteria AIC (Akaike's information criterion) and BIC (Bayesian information criterion) were computed for all 24 models. These four criteria were evaluated together to inform the selection of the most suitable model.

Since all outcome variables were z-transformed, the total variance should always be 1. However, the sums of the estimated level-1 and level-2 variances in Table 100 exceed 1 slightly in most cases. This can be explained by the accompanying, relatively large, standard errors of these estimates. Here, variance partitions were calculated based on the sum of the two estimated variance components, see equation (13) in section 5.6.1.

In general, the variance partition coefficients for the 24 different models range from 9% to 31%, which exceed the 5% threshold suggested above and indicates that multilevel modelling may be beneficial for this data set. Similarly, the variance influence factors range from 2 to 13, which indicates that the effective sample size would be severely overestimated when using single-level models instead.

In particular, students' situational interest varies greatly between groups, almost one third of its variance can be accounted for by level-2 clusters. Here, Figure 36 provides a visualisation of these major differences in students' situational interest across the 28 different groups. This so-called caterpillar plot shows the average situational interest within every group in ascending order together with the respective standard deviation. Here, multi-level models can determine, for example, whether different dispositional interests in physics between level-2 clusters account for this high amount of level-2 variance.

Although interest and self-beliefs were strongly correlated, and both are often interpreted as components of motivation, their variance partition coefficients differed significantly. Indeed, the VPC of students' situational self-beliefs was only about half the size of the VPC of their situational interest. The lowest VPC was found for students' perceived cognitive

[40]The variance inflation factor (VIF) is calculated based on the variance partition coefficient (VPC) and the average size of the level-2 units, see equation 14 in subsection 5.6.1.

activation (9-13%). This could mean that cognitive activation is perceived by all students in a similar way and more or less independent of their prior experiences as a group or the influence of the different tutors in S'Cool LAB.

Table 100. Overview of properties of random intercept models without predictors for 3 different cluster variables (teacher, group, country) and 8 educational outcomes. The listed properties are based on the full sample including extreme values ($N = 509$) and include the model fit criteria AIC & BIC, the estimated amount of level-1 and level-2 variance and their standard errors (SE), variance partition coefficient (VPC) and variance inflation factor (VIF). The best models with the lowest model fit criteria AIC & BIC among the three cluster variables are highlighted in bold.

Educational outcome	Model properties	Cluster variable		
		teacher	group	country
Situational interest IE3	AIC / BIC	1356 / 1365	**1354 / 1362**	1372 / 1380
	Level 1 variance (SE)	0.71 (0.05)	0.74 (0.05)	0.81 (0.05)
	Level 2 variance (SE)	0.31 (0.08)	0.29 (0.09)	0.30 (0.14)
	VPC	30%	28%	27%
	VIF	3	6	11
Curiosity state post CS3	AIC / BIC	**1390 / 1399**	**1390 / 1399**	1398 / 1406
	Level 1 variance (SE)	0.78 (0.05)	0.81 (0.05)	0.85 (0.05)
	Level 2 variance (SE)	0.23 (0.07)	0.22 (0.08)	0.21 (0.10)
	VPC	23%	22%	20%
	VIF	3	5	9
Enjoyment ENJ	AIC / BIC	**1397 / 1405**	**1397 / 1405**	1412 **/ 1404**
	Level 1 variance (SE)	0.83 (0.05)	0.83 (0.05)	0.86 (0.05)
	Level 2 variance (SE)	0.19 (0.07)	0.19 (0.07)	0.22 (0.11)
	VPC	19%	19%	20%
	VIF	3	4	9
Help-fulness HEL	AIC / BIC	**1371 / 1380**	**1373 / 1382**	**1372 / 1380**
	Level 1 variance (SE)	0.74 (0.05)	0.78 (0.05)	0.80 (0.05)
	Level 2 variance (SE)	0.27 (0.08)	0.26 (0.09)	0.36 (0.16)
	VPC	27%	25%	31%
	VIF	3	6	13
Situational self-beliefs SC3	AIC / BIC	**1412 / 1420**	**1413 / 1421**	1421 **/** 1430
	Level 1 variance (SE)	0.84 (0.06)	0.86 (0.06)	0.90 (0.06)
	Level 2 variance (SE)	0.16 (0.05)	0.15 (0.06)	0.16 (0.08)
	VPC	16%	15%	15%
	VIF	2	4	7
Easiness EAS	AIC / BIC	1414 **/** 1422	**1409 / 1418**	1419 **/** 1428
	Level 1 variance (SE)	0.84 (0.06)	0.85 (0.05)	0.90 (0.06)
	Level 2 variance (SE)	0.16 (0.05)	0.15 (0.06)	0.17 (0.09)
	VPC	16%	15%	16%
	VIF	2	4	7
Cognitive activation CA	AIC / BIC	1434 **/** 1443	1434 **/** 1443	**1430 / 1438**
	Level 1 variance (SE)	0.90 (0.06)	0.92 (0.06)	0.92 (0.06)
	Level 2 variance (SE)	0.11 (0.05)	0.09 (0.04)	0.14 (0.08)
	VPC	11%	9%	13%
	VIF	2	3	6
Under-standing UN3	AIC / BIC	1242 / 1250	**1221 / 1229**	1235 / 1243
	Level 1 variance (SE)	0.78 (0.06)	0.77 (0.05)	0.83 (0.06)
	Level 2 variance (SE)	0.23 (0.07)	0.24 (0.08)	0.20 (0.09)
	VPC	23%	24%	19%
	VIF	3	5	8

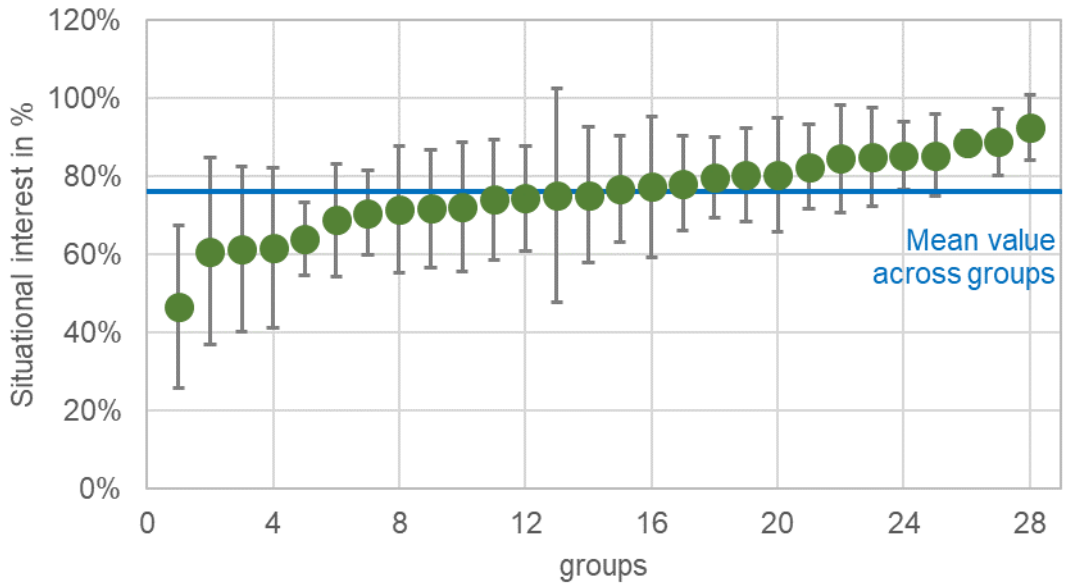


Figure 36. Caterpillar plot: students' situational interest after taking part in the S'Cool LAB hands-on session plotted separately for all 28 groups. Error bars indicate ±1 standard deviation.

Overall, when comparing the three different clustering variables, the clustering according to 'group' leads to the best model fit criteria for seven of the eight educational outcomes. For cognitive activation, clustering according to 'country' leads to slightly better model fit criteria. However, this particular educational outcome showed the lowest VPC and should therefore not be decisive when specifying the common model parameters.

7.8.2 Model specification and additional considerations

The clustering variable 'group' was selected as level-2 unit based on the model fit criteria in Table 100. Here, the variable 'group' refers to the group of students who travelled to CERN together from the same school and took part in the hands-on session in S'Cool LAB together on the same day. Consequently, students are considered to be level-1 units in this two-level model. In many cases, most students in one group had the same physics teacher or were even in the same class. However, sometimes, different classes with different teachers travelled together to reduce costs and organisational efforts. For example, in some cases individual students from different advanced level courses (e.g., chemistry) joined the advanced physics course of their school on their trip to CERN out of interest or because there were places available.

In total, the 509 students, who took part in the main study, came in 28 groups, which, on average, consisted of 19 students. According to the recommendation by Hox (1998), this average cluster size is too small for more complex models, such as cross-level interactions or random slope models. Instead, the data structure is better suited for analysing fixed effects in a random intercept model. However, the accuracy of parameter estimates would

benefit from a larger sample size (see also Table 9). In particular, random intercept models estimate a distinct intercept for every level-2 unit (so-called 'random intercept'), which accounts for the group differences in the average values of the dependent variables. In addition, these types of models allow estimating the effect of predictors on both levels.

In the following, all predictors are considered level-1 predictors, because they were obtained by having individual students fill out questionnaires. Thus, every response reflects a personal view on, for example, novelty experience factors, although the objective novelty experience should be the same for students coming in one group.

Variance components of student-level (level-1) predictors

Before calculating multi-level models on S'Cool LAB's educational outcomes, multi-level modelling was used to estimate by how much these presumed student-level predictors differed systematically between groups. Here, Table 101 lists all predictor variables, which were z-transformed beforehand (i.e., their variance was standardised to 1). Table 101 also lists the amount of variance of predictors that can be accounted for at the student level and at the group level when estimating variance components in a two-level model. Moreover, Figure 37 displays the variance partition coefficients of all novelty factors in decreasing order, which equals the group level variance estimate in this case.

Indeed, some of the variables seemed to be relatively 'pure' student-level variables. For example, 95% of the variance of students' curiosity trait was estimated at the student level, and only 5% at the group level. This indicates that students, who came to CERN together in one group, are not more similar to each other than randomly selected students with respect to their curiosity trait. However, the curiosity trait scale suffered from a ceiling effect. Therefore, the variables' variance was quite small in the first place.

Novelty influence factors with a similar low amount of group-level variance were the dispositional self-beliefs (SC1), students' curiosity state towards particle physics (CS1), students' previous out-of-school science learning experience (EO), and gender. Indeed, since no girls- or boys-schools took part in this study, the gender distribution was not expected to differ largely between groups. However, it is somewhat surprising that students' out-of-school learning experience differs largely for students within one group, because students in one class usually go on school field trips together. Here, individual or family visits to science learning opportunities outside school seem to be more important than school field trips.

Table 101. Estimates of variance components of novelty influence and experience factors that were used as predictors in two-level models on S'Cool LAB's educational outcomes. All variables were z-transformed beforehand. Student-level variance (σ_e^2) and group-level variance estimates (σ_u^2) are reported together with the respective standard errors (SE). Comparatively high amounts of student-level variance (≥ 0.85) are highlighted in blue, comparatively low amounts (≤ 0.65) in orange.

Variables		student-level variance σ_e^2	SE	group-level variance σ_u^2	SE
Novelty influence factors					
IE1	Dispositional physics interest	0.68	0.04	0.36	0.11
SC1	Dispositional physics self-beliefs	0.87	0.06	0.13	0.05
CT	Curiosity trait	0.95	0.06	0.05	0.03
CS1	Curiosity state particle physics	0.85	0.05	0.18	0.07
UN1	Conceptual understanding	0.79	0.06	0.22	0.08
EN	English skills (dichotomised)	0.75	0.05	0.28	0.09
PH	School grade physics	0.75	0.05	0.25	0.08
EH	Experience hands-on experiments	0.42	0.03	0.58	0.17
EO	Experience out-of-school learning	0.94	0.06	0.06	0.03
EP	Experience with particle physics	0.75	0.05	0.29	0.09
SEX	Gender (Femaleness)	0.90	0.06	0.13	0.05
AGE	Age	0.48	0.03	0.48	0.14
Novelty experience factors					
TIN	Tinkering support	0.85	0.05	0.17	0.06
SO	Setting orientation	0.83	0.05	0.19	0.07
PRO	Lack of problems	0.95	0.06	0.05	0.03
CLE	Cognitive load experiments	0.89	0.06	0.10	0.04
TS	Perceived tutor support	0.92	0.06	0.10	0.04

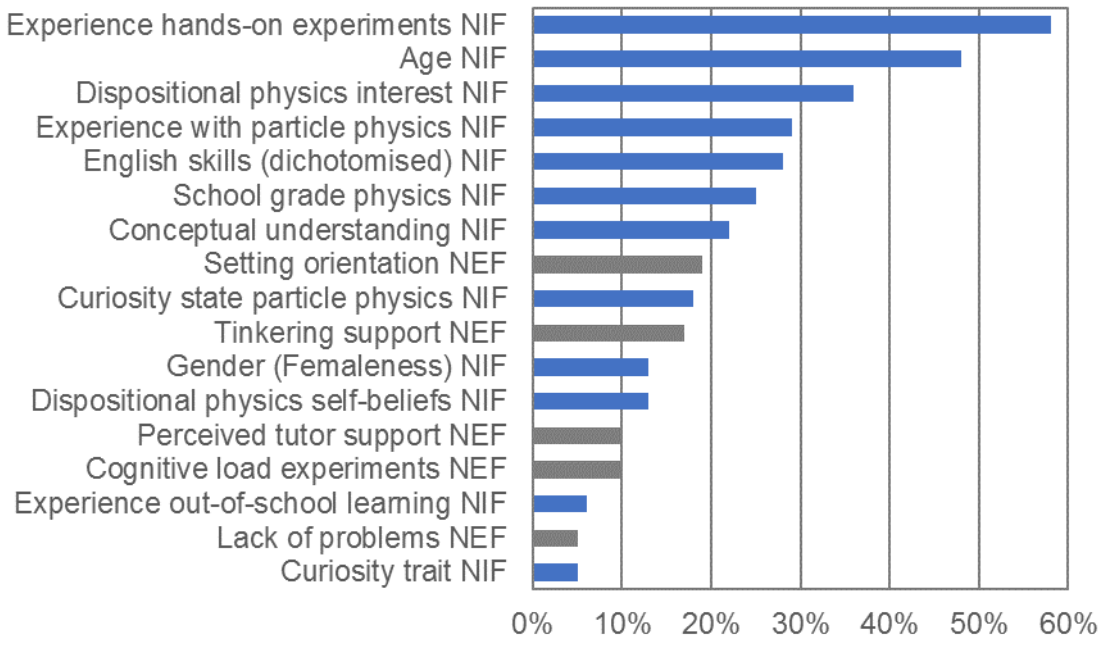


Figure 37. Variance partition coefficients of novelty influence factors (NIFs, blue) and novelty experience factors (NEFs, grey) in decreasing order.

The novelty influence factor 'dispositional interest' showed a significantly larger amount of group-level variance (35%): students, who came to CERN in the same group shared a similar dispositional interest in physics. Therefore, students' dispositional interest should

not act like a 'pure' student-level predictor but instead as a superposition of student- and group-level predictors. Indeed, students in the same group always came from the same school and often shared the same physics teacher. Here, the teachers but also the conditions at the school might affect how much students like their physics lessons.

On the contrary, students' dispositional self-beliefs demonstrated a much smaller amount of group-level variance (13%), although both variables were strongly correlated, and both are often interpreted as components of motivation. This is consistent with the VPCs of students' situational interests and self-beliefs and suggests that self-beliefs are less affected by group-level variables, such as teachers, than other student characteristics.

Interestingly, all novelty experience factors showed relatively low amounts of group-level variance. For example, only 10% of the variance of students' perception of the support by tutors in S'Cool LAB was estimated at the group level. However, tutors differed between groups and it would therefore make sense if this variable had a larger group-level variance. Yet, students' perceptions of tutors were extremely positive and, due to a ceiling effect, the variables' variance was quite small in the first place, which suggests that the differences between tutors were not that important to students. The novelty experience factor 'setting orientation due to preparation', on the other hand, demonstrated a slightly higher proportion of 19% group-level variance. Since the teachers of the group were responsible for preparing their students for the trip to CERN and S'Cool LAB, it makes sense that students within one group perceived their setting orientation similarly. In general, the amount of group-level variance could also be interpreted as an indicator of the mediation role of novelty experience factors: students' perception of the (objective) novelty of an out-of-school learning opportunity may be influenced by their (subjective) interest or previous experiences (novelty influence factors).

As discussed in subsection 7.3.12, students' average experience with hands-on experiments differed significantly between groups. Here, most of the variance was accounted for at the group level and only 42% of the variance was estimated at the student level. Similarly, age was more similar within one group than between students of different groups. Both of these results are not surprising, since students often belonged to the same class, which shared the same classroom experiences and a similar age profile.

In summary, several of the student-level predictors showed relatively high amounts of group-level variance, for example, students' previous experience with hands-on experiments or their dispositional physics interest. Therefore, these student-level

predictors should also reduce group-level variance of S'Cool LAB's educational outcomes when entered in a two-level regression model.

Additional considerations

Ignoring the hierarchical data structure with the cluster variable 'group' would lead to an overestimation of the sample size, on average, by a factor of 5 (see VIFs in Table 100) and in turn to an underestimation of the standard errors for parameter estimates of $\sqrt{5} \approx 2.2$. Therefore, when comparing single-level and two-level models in the following subsections, the p-values for predictors in the single-level models should be smaller than the more accurately estimated p-values in the two-level models. In any case, the Bonferroni correction was applied again to account for the high number of predictors and associated significance tests. This conservative correction was also used for the single-level models, and will, to a certain degree, also compensate for an underestimation of standard error due to the hierarchical data structure.

Before calculating multi-level models, certain assumptions about the data properties needed to be evaluated, for example, the normal distribution of residuals. Indeed, the assumptions of linear multi-level modelling are very similar to those of single-level linear models, see subsection 5.6.1. However, multi-level models are more complex than ordinary linear regression models as they entail the estimation of more parameters, which requires a larger sample size. Here, the sample size of the main study was considered to be sufficiently large for this purpose. In general, assumptions were not evaluated again, as details about the assumptions and their potential violations can be found together with the respective single-level models in section 7.6.

Technically, the screening for uni- and multivariate outliers needs to be repeated for every level and every cluster. However, this procedure is difficult, in particular, for large cluster numbers (Tabachnick & Fidell, 2013). Instead, multi-levels were only computed based on the reduced samples in which certain univariate extreme values had been adjusted and multivariate outliers had been removed based on the results of the single-level models. Appendix 10.8 includes a list of all extreme values and multivariate outliers. To avoid that mere group differences in certain variables lead to the exclusion of extreme values, all extreme values were analysed according to the respective group. Here, no clusters of extreme values in the same variable and group were detected.

Similar to single-level sequential linear models, the estimation of fixed effects in multi-level modelling is robust against violation of normal distribution. However, the standard errors for the regression coefficients can be slightly biased downwards particularly if the number of level-2 units (e.g., groups) is less than 50 (Maas & Hox, 2004). Therefore, the number

of predictors was chosen carefully. In particular, only predictors that were significantly different from zero in the single-level models were selected for the model series. These predictors were then entered stepwise in order of importance as recommended by Raudenbush and Bryk (2002).

7.8.3 MLM: Situational interest

To compare the results of the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.1) was calculated on the outlier-adjusted sample ($N = 493$), see Table 102.

Table 102. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' situational interest: standardised regression coefficients and p-values for the reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), and the model fit criterion R^2 are displayed. Significant predictors[41] are highlighted.

	Single-level full model			Two-level full model		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			0.601			0.847
Gender	0.03		0.274	0.03		0.392
Age	-0.02		0.495	0.01		0.811
Dispositional physics interest IE1	**0.21**	***	8E-07	**0.19**	***	7E-06
Curiosity state particle phys. CS1	**0.20**	***	2E-06	**0.21**	***	4E-07
Experience hands-on experiments EH	-0.03		0.277	0.01		0.869
English skills (binary) EN	0.04		0.185	0.02		0.531
Experience with OSLEOs EO	0.01		0.777	0.02		0.605
Experience with particle physics EP	-0.04		0.265	-0.02		0.611
Tinkering support (novelty) TIN	**0.29**	***	3E-10	**0.26**	***	3E-09
Setting orientation (novelty) SO	0.07		0.074	0.07		0.064
Lack of problems (novelty) PRO	-0.07		0.077	-0.05		0.128
Cognitive load experiments CLE	-0.04		0.321	-0.07		0.071
Tutor support TS	**0.27**	***	3E-11	**0.27**	***	1E-10
Model fit:	Adjusted $R^2 = 0.55$			Analogue $R^2 = 0.58$		

The size of significant regression coefficients was similar to the results of the single-level sequential regression. However, p-values were slightly larger in the two-level model, which was expected due to the smaller effective sample size caused by the clustering of the data. As in the single-level model, only four of the 13 predictors showed regression coefficients significantly different from zero with $p < 0.004$ (Bonferroni correction with $m = 13$). Specifically, the four predictors are the novelty influence factors 'dispositional interest' and 'curiosity state', and the novelty experience factors 'tinkering support' and 'tutor support'.

[41] The Bonferroni correction reduces the required p-values with $m = 13$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

In total, all 13 predictors in the two-level model accounted for approximately 58% of the variance in students' situational interest. However, this R^2 value should be treated cautiously as it does not penalise the high number of predictors. Hence, it is also slightly larger than the adjusted amount of variance accounted for by the predictors in the single-level model (55%).

To understand better which of the significant predictors reduced variance at the group level and the student level, a series of multi-level models was calculated in which the four significant predictors were entered step by step. These four models were then compared with an empty model without predictors and the full model with all predictors. In particular, the following six models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: dispositional interest
- **Model 2:** 2 predictors: dispositional interest, curiosity state
- **Model 3:** 3 predictors: dispositional interest, curiosity state & tinkering support
- **Model 4:** 4 predictors: dispositional interest, curiosity state, tinkering support & tutor support
- **Full model for comparison:** All 13 predictors

Table 103 summarises the different models and their properties. When evaluating the model fit parameters AIC and BIC, entering predictors improved the model compared to the intercept-only model 0. Specifically, every predictor improved the model compared to the previous step. However, the best model fit was reached once the four significant predictors were included. Here, model 4 accounted for 57% of the variance of the dependent variable. Adding the other 9 predictors increased the model fit parameters AIC and BIC and only accounted for 1% of the variance of the dependent variable.

In the empty model 0, the group-level variance was estimated with 28% while the student-level variance was estimated with 72%[42]. Theoretically, including 'pure' student-level predictors should only affect the estimated student-level variance, and group-level predictors should only affect group-level variance. However, in model 4 with four student-level predictors, only 24% of the original estimate for the group-level variance of the empty

[42] Since all outcome variables were z-transformed, the total variance should be 1. However, the sum of the estimated student and group-level variance in Table 103 exceeds 1 slightly ($0.74 + 0.29 = 1.03$). This can be explained when considering the size of the standard errors of the estimates but also because outliers were removed after the z-transformation. Therefore, variance partitions were calculated based on the sum of the two estimated variance components.

model remains. Therefore, the four student-level predictors reduced both the student- and group-level variance.

Table 103. Set of multi-level linear models on students' situational interest (sample without outliers $N = 493$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors $(\sigma^2,\ SE)$. The models were compared using the model fit criteria analogue R^2, AIC and BIC.

Model	0	1		2		3		4		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	
Intercept	0.01	0.00	0.986	0.00	0.991	0.01	0.895	0.01	0.863	
Disp. Interest		**0.44**	3E-24	**0.24**	1E-06	**0.22**	2E-07	**0.21**	2E-07	**See Table 102**
Curiosity state				**0.33**	2E-11	**0.21**	1E-06	**0.21**	1E-07	
Tinker support						**0.45**	6E-36	**0.29**	5E-13	
Tutor support								**0.28**	4E-12	
Random part										
student level σ_e^2 (SE)	0.74 (0.05)	0.59 (0.04)		0.53 (0.04)		0.39 (0.03)		0.36 (0.02)		0.35 (0.02)
group level σ_u^2 (SE)	0.29 (0.09)	0.12 (0.05)		0.12 (0.04)		0.07 (0.03)		0.07 (0.02)		0.07 (0.03)
$Analogue\ R^2$		0.29		0.35		0.54		**0.57**		0.58
AIC	1354	1187		1146		994		**950**		984
BIC	1362	1195		1154		1002		**959**		993

In model 1, after entering students' dispositional interest as predictor, the student-level variance decreased from 0.74 to 0.59, while, at the same time, the group-level variance decreased from 0.29 to 0.12. As expected from the results displayed in Table 101, including dispositional interest affected variance on both levels since this variable can be interpreted as a superposition of a student- and group-level predictor.

In model 2, students' pre-visit curiosity was entered as second predictor, which further reduced the student-level variance slightly to 0.53 but did not further reduce the group-level variance. After controlling for students' dispositional interest, curiosity state seemed to be a relatively 'pure' student-level variable, which is consistent with the variance components of this predictor, see Table 101.

In model 3, students' perceived 'tinkering support' was entered as third predictor, which further reduced the student-level variance and, to a small degree, the group-level variance. This suggests that perceived tinkering support is indeed a superposition of the (objective) tinkering support provided in S'Cool LAB on the day of the visit and students' subjective perception of this support. Here, the subjective component had a stronger influence.

Last, in model 4, entering students' perceived 'tutor support' only reduced the student-level variance slightly, which is consistent with the variance components of this predictor, see Table 101. However, due to the strong correlation between the variables tinkering support

and tutor support, it is not trivial to interpret this effect separately because the order in which predictors were entered affected the respective effects on the variance components.

In summary, the two novelty influence factors (dispositional interest & curiosity state) and the two novelty experience factors (tinkering support & tutor support) significantly improved the multi-level model on students' situational interest. Together, they accounted for more than half of the variability in the dependent variable. In particular, they reduced the student-level variance from 0.74 to 0.36 and the group-level variance from 0.29 to 0.07. Here, especially the novelty influence factors reduced the relevance of the random intercept term in this model. **Indeed, due to the largely reduced group-level variance after adding significant predictors, on could argue that multi-level modelling is not necessary for this data set in this case.**

7.8.4 MLM: Curiosity state particle physics

To compare the results from the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.2) was calculated on the outlier-adjusted sample ($N = 485$), see Table 104.

Table 104. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' curiosity state: standardised regression coefficients and p-values for reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), and the model fit criterion R^2 are displayed. Significant predictors[43] are highlighted.

	Single-level full model			two-level full model		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			0.127	-0.05		0.442
Curiosity trait CT	0.09		0.014	0.09		0.025
Curiosity state particle phys. CS1	**0.40**	***	5E-17	**0.43**	***	2E-18
Dispositional physics interest IE1	0.08		0.065	0.08		0.082
Experience hands-on experiments EH	-0.01		0.762	-0.02		0.636
English skills (binary) EN	0.06		0.080	0.08		0.223
Experience with OSLEOs EO	-0.01		0.729	-0.01		0.795
Experience with particle physics EP	-0.03		0.327	-0.04		0.225
Tinkering support (novelty) TIN	**0.16**	*	1E-03	0.13	ns	0.006
Setting orientation (novelty) SO	0.02		0.688	0.00		0.905
Lack of problems (novelty) PRO	-0.05		0.193	-0.03		0.467
Cognitive load experiments CLE	0.01		0.903	0.00		0.982
Tutor support TS	**0.22**	***	3E-07	**0.24**	***	3E-07
Model fit:	$Adjusted\ R^2 = 0.49$			$Analogue\ R^2 = 0.54$		

[43] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

As expected, the size of significant regression coefficients and the amount of modelled variance were similar to the results of the single-level sequential regression analysis, but p-values were slightly larger for some of the predictors. As before, the novelty influence factor 'curiosity state', and the novelty experience factor 'tutor support' showed regression coefficients significantly different from zero with $p < 0.004$ (Bonferroni correction with $m = 12$). However, due to an increased p-value, the novelty experience factor 'tinkering support' was no longer significant when applying the Bonferroni correction. In total, all 12 predictors in the two-level model accounted for approximately 54% of the variance in students' curiosity towards particle physics after taking part in the hands-on session.

To understand better, which of the significant predictors reduced variance at the group level and the student level, a series of multi-level models was calculated in which the significant predictors were entered step by step. Here, the predictor 'tinkering support' was considered, even though its p-value was problematic. These models were then compared with an empty model without predictors and the full model with all predictors. In particular, the following five models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: curiosity state
- **Model 2:** 2 predictors: curiosity state & tinkering support
- **Model 3:** 3 predictors: curiosity state, tinkering support & tutor support
- **Full model for comparison:** All 12 predictors

Table 105 summarises the different models and their properties.

Table 105. Set of multi-level linear models on students' curiosity state (sample without outliers $N = 485$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors (σ^2, SE). The models were compared using the model fit criteria analogue R^2, AIC and BIC.

Model	0	1		2		3		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	
Intercept	0.02	0.01	0.850	0.01	0.855	0.00	0.932	**See Table 104**
Curiosity state CS1		**0.60**	3E-46	**0.51**	6E-36	**0.51**	3E-37	
Tinkering support TIN				**0.27**	1E-13	**0.13**	0.002	
Tutor support TS						**0.24**	2E-07	
Random part								
student level σ_e^2 (SE)	0.73 (0.05)	0.49 (0.03)		0.45 (0.03)		0.42 (0.03)		0.42 (0.03)
group level σ_u^2 (SE)	0.19 (0.07)	0.08 (0.03)		0.04 (0.02)		0.04 (0.02)		0.04 (0.02)
$Analogue\ R^2$		0.43		0.51		**0.54**		0.54
AIC	1278	1080		1031		**1008**		1035
BIC	1286	1088		1039		**1016**		1043

When evaluating the model fit parameters AIC and BIC, entering predictors improved the model compared to the intercept-only model 0. Specifically, every predictor improved the model compared to the previous step. However, the best model fit was reached once the three strongest predictors were included. Here, model 3 accounted for 54% of the variance of the dependent variable. Adding the other 9 predictors increased the model fit parameters AIC and BIC and did not account for additional variance of the dependent variable.

In summary, the novelty influence factor (curiosity state) and the two novelty experience factors (tinkering & tutor support) significantly improved this multi-level model. Together, they accounted for more than half of the variability in the dependent variable. In particular, they reduced the student-level variance from 0.73 to 0.42 and the group-level variance from 0.19 to 0.04, which is consistent with the variance components of these predictors, see Table 101. The model series in Table 105 also shows that the p-value of the predictor 'tinkering support' is reduced significantly when entering the (correlated) predictor 'tutor support'. This suggests the influence of multicollinearity issues, which justifies interpreting 'tinkering support' as a strong predictor nonetheless.

7.8.5 MLM: Perceived enjoyment

To compare the results of the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.3) was calculated on the outlier-adjusted sample $(N = 492)$, see Table 106. As before, the size of significant regression coefficients and the amount of modelled variance were similar to the results of the single-level sequential regression analysis, but p-values were slightly larger.

As in the single-level model, the novelty influence factor 'dispositional interest', and the novelty experience factors 'tinkering support' and 'tutor support' showed regression coefficients significantly different from zero with $p < 0.005$ (Bonferroni correction with $m = 10$). In total, all 10 predictors in the two-level model accounted for approximately 54% of the variance in students' enjoyment.

To understand better, which of the significant predictors reduced variance at the group level and at the student level, a series of multi-level models was calculated in which the significant predictors were entered step by step. These models were then compared with an empty model without predictors and the full model with all predictors.

Table 106. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' perceived enjoyment: standardised regression coefficients and p-values for reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), and the model fit criterion R^2 are displayed. Significant predictors[44] are highlighted.

	Single-level full model			two-level full model		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			0.298	0.04		0.540
Dispositional physics interest IE1	**0.12**	**	9E-04	**0.11**	*	0.002
Experience hands-on experiments EH	-0.04		0.226	-0.03		0.410
English skills (binary) EN	-0.04		0.259	-0.07		0.266
Experience with OSLEOs EO	-0.02		0.465	-0.01		0.755
Experience with particle physics EP	-0.04		0.262	-0.02		0.609
Tinkering support (novelty) TIN	**0.37**	***	2E-14	**0.31**	***	3E-12
Setting orientation (novelty) SO	0.05		0.219	0.05		0.153
Lack of problems (novelty) PRO	-0.04		0.306	-0.02		0.672
Cognitive load experiments CLE	-0.05		0.267	-0.07		0.104
Tutor support TS	**0.33**	***	4E-14	**0.34**	***	7E-14
Model fit:	$Adjusted\ R^2 = 0.48$			$Analogue\ R^2 = 0.54$		

In particular, the following four models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: dispositional interest
- **Model 2:** 2 predictors: dispositional interest & tinkering support
- **Model 3:** 3 predictors: dispositional interest, tinkering support & tutor support
- **Full model for comparison:** All 10 predictors

Table 107 summarises the different models and their properties. Entering significant predictors improved the model compared to the intercept-only model 0. Specifically, every predictor improved the model compared to the previous step. However, the best model fit was reached once the three significant predictors were included. Here, model 3 accounted for 54% of the variance of the dependent variable. Adding the other 7 predictors increased the model fit parameters AIC and BIC and did not account for additional variance of the dependent variable.

In model 1, after entering students' dispositional interest as predictor, the student-level variance decreased from 0.72 to 0.68, while, at the same time, the group-level variance decreased from 0.17 to 0.12. As expected from the results displayed in Table 101, including the predictor 'dispositional interest' affected variance on both levels. However, for the

[44] The Bonferroni correction reduces the required p-values with $m = 10$ to
* $p < 0.005$, ** $p < 0.001$ and *** $p < 0.0001$

outcome variable 'enjoyment', this predictor even had a stronger effect on the group-level variance.

In summary, the novelty influence factor (curiosity state) and the two novelty experience factors (tinkering support & tutor support) significantly improved this multi-level model. Together, they accounted for more than half of the variability in the dependent variable. In particular, they reduced the student-level variance from 0.73 to 0.42 and the group-level variance from 0.19 to 0.04, which is consistent with the variance components of the predictors, see Table 101.

Table 107. Set of multi-level linear models on students' perceived enjoyment (sample without outliers $N = 492$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors (σ^2, SE). The models were compared using the model fit criteria analogue R^2, AIC and BIC.

Model	0	1		2		3		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	
Intercept	0.00	0.00	0.961	0.00	0.950	0.00	0.947	**See Table 106**
Dispositional interest IE1		**0.27**	1E-09	**0.14**	9E-05	**0.13**	2E-04	
Tinkering support TIN				**0.54**	5E-45	**0.34**	1E-16	
Tutor support TS						**0.35**	4E-15	
Random part								
student level σ_e^2 (SE)	0.72 (0.05)	0.68 (0.04)		0.46 (0.03)		0.41 (0.03)		0.41 (0.03)
group level σ_u^2 (SE)	0.17 (0.06)	0.12 (0.04)		0.05 (0.02)		0.05 (0.02)		0.05 (0.02)
$Analogue\ R^2$		0.20		0.49		**0.54**		0.54
AIC	1287	1255		1063		**1006**		1030
BIC	1295	1264		1072		**1014**		1038

7.8.6 MLM: Perceived helpfulness

To compare the results of the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.3) was calculated on the outlier-adjusted sample $(N = 493)$, see Table 108. As before, the size of significant regression coefficients and the amount of modelled variance were similar to the results from the single-level sequential regression analysis, but p-values were slightly larger.

As in the single-level model, the novelty influence factor 'dispositional interest', and the novelty experience factors 'tinkering support' and 'tutor support' showed regression coefficients significantly different from zero with $p < 0.005$ (Bonferroni correction with $m = 10$). However, due to an increased p-value, the novelty experience factor 'lack of problems' was no longer significant when applying the conservative Bonferroni correction. In any case, as discussed in subsection 7.7.3, this predictor should be interpreted with caution because its regression coefficient was caused by a suppression effect in the single-

level model. In total, all 10 predictors in the two-level model accounted for approximately 39% of the variance in students' perceived helpfulness.

Table 108. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' perceived helpfulness: standardised regression coefficients and p-values for reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), and the model fit criterion R^2 are displayed. Significant predictors[45] are highlighted.

	Single-level			two-level		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			6E-10	0.07		0.404
Dispositional physics interest IE1	**0.16**	***	5E-05	**0.14**	*	0.001
Experience hands-on experiments EH	0.07		0.066	0.05		0.311
English skills (binary) EN	-0.02		0.563	-0.12		0.114
Experience with OSLEOs EO	0.00		0.946	0.02		0.675
Experience with particle physics EP	-0.07		0.047	-0.10		0.009
Tinkering support (novelty) TIN	**0.36**	***	8E-12	**0.30**	***	2E-09
Setting orientation (novelty) SO	0.09		0.041	0.10		0.022
Lack of problems (novelty) PRO	-0.16	**	5E-04	-0.11		0.015
Cognitive load experiments CLE	0.01		0.781	0.01		0.871
Tutor support TS	**0.27**	***	8E-09	**0.26**	***	2E-07
Model fit:	$Adjusted\ R^2 = 0.39$			$Analogue\ R^2 = 0.39$		

To understand better, which of the significant predictors reduces variance at the group level and art the student level, a series of multi-level models was calculated in which the significant predictors were entered step by step. Here, the predictor 'lack of problems' was considered, even though its p-value was problematic, to compare its behaviour in a two-level model with the suppression effect in the single-level model. These models were then compared with an empty model without predictors and the full model with all predictors. In particular, the following six models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: dispositional interest
- **Model 2:** 2 predictors: dispositional interest & tinkering support
- **Model 3:** 3 predictors: dispositional interest, tinkering support & lack of problems
- **Model 4:** 4 predictors: dispositional interest, tinkering support, lack of problems & tutor support
- **Full model for comparison:** All 10 predictors

Table 109 summarises the different models and their properties. When evaluating the model fit parameters AIC and BIC, entering predictors improved the model compared to

[45] The Bonferroni correction reduces the required p-values with $m = 10$ to
* $p < 0.005$, ** $p < 0.001$ and *** $p < 0.0001$

the intercept-only model 0. Specifically, every predictor, except for 'lack of problems', improved the model compared to the previous step. The best model fit was reached once the strongest predictors were included. Here, model 4 accounts for 38% of the variance of the dependent variable. Adding the other 6 predictors increased the model fit parameters AIC and BIC and did not account for significant additional variance of the dependent variable.

In model 1, after entering students' dispositional interest as predictor, the student-level variance decreased from 0.73 to 0.69, while, at the same time, the group-level variance decreased from 0.24 to 0.18. As expected from the results displayed in Table 101, including the predictor 'dispositional interest' affected variance on both levels. As in the two-level model on the outcome variable 'enjoyment', this predictor even had a stronger effect on the group-level variance.

As suspected from the results of the single-level model, the regression coefficient of the predictor 'lack of problems' is caused by suppressor variables. Indeed, the size of its regression coefficient increased from model 3 to 4, while its p-value decreased without improving the model fit criteria AIC and BIC. Hence, this predictor should not be interpreted.

Table 109. Set of multi-level linear models on students' perceived helpfulness (sample without outliers $N = 493$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors (σ^2, SE). The models were compared using the model fit criteria analogue R^2, AIC and BIC. A suppressor effect is highlighted in red.

Model	0	1		2		3		4		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	
Intercept	-0.01	-0.01	0.941	-0.01	0.899	-0.01	0.936	-0.01	0.904	**See Table** 108
Disp. Interest		**0.26**	1E-08	**0.16**	8E-05	**0.17**	6E-05	**0.15**	1E-04	
Tinkering support				**0.43**	1E-26	**0.46**	4E-27	**0.33**	9E-12	
Lack of problems						-0.08	0.029	-0.11	0.003	
Tutor support								**0.27**	9E-08	
Random part										
student level σ_e^2 (SE)	0.73 (0.05)	0.69 (0.05)		0.56 (0.04)		0.56 (0.04)		0.52 (0.03)		0.51 (0.03)
group level σ_u^2 (SE)	0.24 (0.08)	0.18 (0.06)		0.12 (0.04)		0.10 (0.04)		0.10 (0.04)		0.10 (0.04)
$Analogue\ R^2$		0.13		0.32		0.34		**0.38**		0.39
AIC	1303	1275		1166		1166		**1141**		1151
BIC	1311	1284		1174		1174		**1150**		1159

In summary, the novelty influence factor (dispositional interest) and the two novelty experience factors (tinkering support & tutor support) significantly improved this multi-level model. Together, they accounted for approximately one third of the variability in the dependent variable. In particular, they reduced the student-level variance from 0.73 to 0.52

and the group-level variance from 0.24 to 0.10, which is consistent with the variance components of these predictors, see Table 101.

7.8.7 MLM: Situational self-beliefs

To compare the results of the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.4) was calculated on the outlier-adjusted sample $(N = 495)$, see Table 110. As before, the size of significant regression coefficients and the amount of modelled variance were similar to the results from the single-level sequential regression analysis, but p-values were slightly larger for some of the predictors.

As in the single-level model, the novelty influence factor 'dispositional self-beliefs', and the novelty experience factors 'setting orientation', 'cognitive load' and 'tutor support' showed regression coefficients significantly different from zero with $p < 0.004$ (Bonferroni correction with $m = 12$). In total, all 12 predictors in the wo-level model accounted for approximately 58% of the variance in students' situational self-beliefs.

Table 110. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' situational self-beliefs: standardised regression coefficients and p-values for reduced sample without outliers $(\beta_{w/o}, p_{w/o})$, and the model fit criterion R^2 are displayed. Significant predictors[46] are highlighted.

	Single-level full model			two-level full model		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			0.021	-0.03		0.635
Gender	0.00		0.963	-0.02		0.722
Age	0.02		0.466	0.02		0.615
Disp. self-beliefs interest SC1	**0.21**	***	2E-09	**0.19**	***	3E-08
Experience hands-on experiments EH	0.00		0.958	0.00		0.972
English skills (binary) EN	0.06		0.055	0.07		0.283
Experience with OSLEOs EO	0.03		0.285	0.05		0.280
Experience with particle phys. EP	-0.02		0.503	-0.01		0.712
Tinkering support (novelty) TIN	0.10		0.036	0.08		0.074
Setting orientation (novelty) SO	**0.25**	***	4E-10	**0.23**	***	7E-10
Lack of problems (novelty) PRO	0.02		0.570	0.03		0.385
Cognitive load experiments CLE	**-0.31**	***	7E-13	**-0.31**	***	9E-14
Tutor support TS	**0.16**	***	1E-04	**0.19**	***	2E-05
Model fit:	$Adjusted\ R^2 = 0.52$			$Analogue\ R^2 = 0.58$		

[46] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

To understand better, which of the significant predictors reduced variance at the group level and at the student level, a series of multi-level models was calculated in which the significant predictors were entered step by step. These models were then compared with an empty model without predictors and the full model with all predictors. In particular, the following six models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: dispositional self-beliefs
- **Model 2:** 2 predictors: dispositional self-beliefs & setting orientation
- **Model 3:** 3 predictors: dispositional self-beliefs, setting orientation & cognitive load
- **Model 4:** 4 predictors: dispositional self-beliefs, setting orientation, cognitive load & tutor support
- **Full model for comparison:** All 12 predictors

Table 111 summarises the different models and their properties. Entering significant predictors improved the model compared to the intercept-only model 0. Specifically, every predictor improved the model compared to the previous step. However, the best model fit was reached once the four significant predictors were included. Here, model 4 accounted for 58% of the variance of the dependent variable. Adding the other 6 predictors increased the model fit parameters AIC and BIC and did not account for additional variance of the dependent variable.

Table 111. Set of multi-level linear models on students' situational self-beliefs (sample without outliers $N = 495$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors (σ^2, SE). The models were compared using the model fit criteria analogue R^2, AIC and BIC.

Model	0	1		2		3		4		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	
Intercept	0.01	0.02	0.784	0.02	0.691	0.01	0.885	0.00	0.972	
Disp. self-beliefs		**0.36**	1E-17	**0.25**	3E-11	**0.18**	6E-08	**0.19**	3E-09	**See Table 110**
Setting orientation				**0.43**	3E-27	**0.31**	4E-18	**0.25**	2E-12	
Cognitive load						**-0.41**	3E-29	**-0.35**	2E-22	
Tutor support								**0.23**	7E-10	
Random part										
student level σ_e^2 (SE)	0.78 (0.05)	0.69 (0.06)		0.56 (0.04)		0.43 (0.03)		0.39 (0.03)		0.39 (0.03)
group level σ_u^2 (SE)	0.11 (0.04)	0.06 (0.03)		0.03 (0.01)		0.03 (0.01)		0.03 (0.01)		0.03 (0.01)
$Analogue\ R^2$		0.25		0.41		0.54		**0.58**		0.58
AIC	1323	1256		1145		1024		**991**		1019
BIC	1331	1264		1153		1032		**999**		1028

In model 1, after entering students' dispositional self-beliefs as predictor, the student-level variance decreased from 0.78 to 0.69, while, at the same time, the group-level variance

decreased from 0.11 to 0.06. As expected from the results displayed in Table 101, including the predictor 'dispositional self-beliefs' affected variance on both levels.

In model 2, the novelty experience factor 'setting orientation' was entered as second predictor, which further reduced the student-level variance from 0.60 to 0.56 and also slightly reduced the group-level variance from 0.06 to 0.03. However, adding more novelty experience factors in models 3 and 4 did not reduce the group-level variance further. These variables only affected the student-level variance, which is consistent with the variance components of this predictor, see Table 101. In particular, all novelty experience factors displayed relatively small group-level variance components with 'setting orientation' having the largest VPC of 18%.

In summary, the novelty influence factor (dispositional self-beliefs) and the three novelty experience factors (setting orientation, cognitive load and tutor support) significantly improved this multi-level model. Together, they accounted for more than half of the variability in the dependent variable. In particular, they reduced the student-level variance from 0.78 to 0.39 and the group-level variance from 0.11 to 0.03, which is consistent with the variance components of the predictors, see Table 101.

As discussed before, the educational outcome 'situational self-beliefs' and also the predictor 'dispositional self-beliefs' displayed a relatively small variance partition coefficient in the first place. This came as a surprise because it would make sense that students' self-beliefs are affected by the school and teacher in a similar way as their physics interest. However, based on the results of this study, it is not clear how these differences can be interpreted.

7.8.8 MLM: Perceived easiness

To compare the results of the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.5) was calculated on the outlier-adjusted sample $(N = 493)$, see Table 112.

As before, the size of significant regression coefficients and the amount of modelled variance were similar to the results from the single-level sequential regression analysis, but p-values were slightly larger for most of the predictors. As in the single-level model, only the novelty experience factors 'tinkering support', 'setting orientation', 'lack of problems' and 'cognitive load' showed regression coefficients significantly different from zero with $p < 0.005$ (Bonferroni correction with $m = 10$). In total, all 10 predictors in the two-level model accounted for approx. 53% of the variance in perceived easiness.

Table 112. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' perceived easiness: standardised regression coefficients and p-values for reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), and the model fit criterion R^2 are displayed. Significant predictors[47] are highlighted.

	Single-level			two-level		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			5E-19	0.01		0.814
Disp. self-beliefs interest SC1	-0.05		0.171	-0.05		0.130
Experience hands-on experim. EH	0.01		0.760	-0.01		0.893
English skills (binary) EN	0.00		0.966	-0.02		0.740
Experience with OSLEOs EO	0.04		0.196	0.04		0.381
Experience with particle phys. EP	0.04		0.170	0.03		0.311
Tinkering support (novelty) TIN	**0.25**	***	4E-08	**0.25**	***	9E-08
Setting orientation (novelty) SO	**0.17**	***	6E-06	**0.17**	***	6E-06
Lack of problems (novelty) PRO	**0.14**	**	4E-04	**0.13**	*	0.001
Cognitive load experiments CLE	**-0.36**	***	2E-16	**-0.36**	***	5E-16
Tutor support TS	0.02		0.601	0.05		0.272
Model fit:	$Adjusted\ R^2 = 0.52$			$Analogue\ R^2 = 0.53$		

To understand better, which of the significant predictors reduced variance at the group level and at the student level, a series of multi-level models was calculated in which the significant predictors were entered step by step. These models were then compared with an empty model without predictors and the full model with all predictors. In particular, the following six models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: tinkering support
- **Model 2:** 2 predictors: tinkering support & setting orientation
- **Model 3:** 3 predictors: tinkering support, setting orientation & lack of problems
- **Model 4:** 4 predictors: tinkering support, setting orientation, lack of problems & cognitive load
- **Full model for comparison:** All 10 predictors

Table 113 summarises the different models and their properties.

Entering significant predictors improved the model compared to the intercept-only model 0. Specifically, every predictor improved the model compared to the previous step. However, the best model fit was reached once the four significant predictors were included. Here, model 4 accounted for 54% of the variance of the dependent variable. Adding the other 6

[47] The Bonferroni correction reduces the required p-values with $m = 10$ to
* $p < 0.005$, ** $p < 0.001$ and *** $p < 0.0001$

predictors increased the model fit parameters AIC and BIC and decreased the modelled variance of the dependent variable.

Table 113. Set of multi-level linear models on students' perceived easiness (sample without outliers $N = 493$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors $(\sigma^2,\ SE)$. The models were compared using the model fit criteria analogue R^2, AIC and BIC.

Model	0	1		2		3		4		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	
Intercept	0.01	0.01	0.835	0.02	0.783	0.01	0.912	0.00	0.975	**See Table 112**
Tinkering support		**0.56**	6E-41	**0.46**	6E-26	**0.35**	2E-16	**0.27**	2E-11	
Setting orientation				**0.23**	6E-08	**0.22**	2E-08	**0.17**	3E-06	
Lack of problems						**0.30**	8E-16	**0.13**	8E-04	
cognitive load								**-0.36**	2E-16	
Random part										
student level σ_e^2 (SE)	0.84 (0.06)	0.60 (0.04)		0.56 (0.04)		0.50 (0.03)		0.44 (0.03)		0.44 (0.03)
group level σ_u^2 (SE)	0.15 (0.05)	0.07 (0.03)		0.06 (0.03)		0.04 (0.02)		0.02 (0.01)		0.03 (0.02)
$Analogue\ R^2$		0.33		0.38		0.46		**0.54**		0.53
AIC	1362	1187		1163		1103		**1040**		1062
BIC	1370	1196		1171		1112		**1049**		1070

In summary, the four novelty experience factors significantly improved this multi-level model. Together, they accounted for more than half of the variability in the dependent variable. In particular, they reduced mostly the student-level variance from 0.84 to 0.44 but also the group-level variance from 0.15 to 0.02, which is consistent with the variance components of the predictors, see Table 101. Here, the educational outcome 'perceived easiness' had already displayed a relatively small variance partition coefficient in the first place but the group-level variance disappeared almost completely when including predictors.

7.8.9 MLM: Perceived cognitive activation

To compare the results of the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.6) was calculated on the outlier-adjusted sample $(N = 392)$, see Table 114. As before, the size of significant regression coefficients and the amount of modelled variance were similar to the results from the single-level sequential regression analysis, but p-values were slightly larger for some of the predictors.

As in the single-level model, the novelty experience factors 'tinkering support', 'setting orientation' and 'tutor support' showed regression coefficients significantly different from zero with $p < 0.004$ (Bonferroni correction with $m = 13$). In total, all 13 predictors in the

two-level model accounted for approximately 55% of the variance in students' perceived cognitive activation.

Table 114. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' situational interest: standardised regression coefficients and p-values for reduced sample without outliers ($\beta_{w/o}, p_{w/o}$), and the model fit criterion R^2 are displayed. Significant predictors[48] are highlighted.

	Single-level full model			two-level full model		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			0.208	-0.05		0.468
Pre-visit understanding score UN1	-0.05		0.208	-0.03		0.476
Physics grade PH	0.08		0.091	0.06		0.191
Dispositional physics interest IE1	0.03		0.529	0.05		0.297
Dispositional self-beliefs SC1	0.14		0.009	0.12		0.025
Experience hands-on experiments EH	-0.02		0.607	0.00		0.979
English skills (binary) EN	0.06		0.145	0.08		0.344
Experience with OSLEOs EO	0.07		0.056	0.08		0.080
Experience with particle physics EP	-0.09		0.013	-0.06		0.123
Tinkering support (novelty) TIN	**0.32**	***	2E-09	**0.33**	***	1E-09
Setting orientation (novelty) SO	**0.22**	***	7E-07	**0.23**	***	2E-07
Lack of problems (novelty) PRO	-0.07		0.120	-0.07		0.135
Cognitive load experiments CLE	-0.04		0.437	-0.05		0.340
Tutor support TS	**0.26**	***	3E-08	**0.29**	***	6E-08
Model fit:	$Adjusted\ R^2 = 0.53$			$Analogue\ R^2 = 0.55$		

To understand better, which of the significant predictors reduced variance at the group level and at the student level, a series of multi-level models was calculated in which the significant predictors were entered step by step. These models were then compared with an empty model without predictors and the full model with all predictors. In particular, the following five models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: tinkering support
- **Model 2:** 2 predictors: tinkering support & setting orientation
- **Model 3:** 3 predictors: tinkering support, setting orientation & tutor support
- **Full model for comparison:** All 13 predictors

Table 115 summarises the different models and their properties.

[48] The Bonferroni correction reduces the required p-values with $m = 13$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

Entering significant predictors improved the model compared to the intercept-only model 0. Specifically, every predictor improved the model compared to the previous step. Contrary to the other multi-level models described in subsections 7.8.3 to 7.8.8, the best model fit was only reached once all 13 predictors were included in the full model. Here, the full model accounted for 55% of the variance of the dependent variable. In model 1, after entering 'tinkering support' as predictor, the student-level variance decreased from 0.87 to 0.59. As expected from the results displayed in Table 101, this predictor acted mainly on the student-level variance. Nevertheless, it also slightly reduced the group-level variance from 0.07 to 0.02. However, adding more novelty experience factors in models 2 and 3 did not reduce the group-level variance any further. These variables only affected the student-level variance.

Table 115. Set of multi-level linear models on students' perceived cognitive activation (sample without outliers $N = 392$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors (σ^2, SE). The models were compared using the model fit criteria analogue R^2, AIC and BIC.

Model	0	1		2		3		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	
Intercept	-0.01	-0.01	0.863	-0.01	0.870	-0.01	0.860	**See Table 115**
Tinkering support		**0.58**	2E-46	**0.44**	3E-27	**0.32**	1E-12	
Setting orientation				**0.31**	1E-14	**0.30**	1E-14	
Tutor support						**0.23**	2E-08	
Random part								
student level σ_e^2 (SE)	0.87 (0.06)	0.59 (0.04)		0.52 (0.03)		0.48 (0.03)		0.42 (0.03)
group level σ_u^2 (SE)	0.07 (0.04)	0.02 (0.01)		0.03 (0.02)		0.03 (0.02)		0.03 (0.02)
$Analogue\ R^2$		0.39		0.45		0.49		**0.55**
AIC	1375	1173		1119		1092		**841**
BIC	1383	1182		1128		1101		**849**

In summary, the three novelty experience factors (tinkering support, tutor support and setting orientation) significantly improved this multi-level model. Together, they accounted for half of the variability in the dependent variable. The outcome 'cognitive activation' displayed the smallest VPC among all educational outcomes. Nevertheless, the predictors reduced not only the student-level variance from 0.87 to 0.48, but also, to a smaller degree, the group-level variance from 0.07 to 0.03.

7.8.10 MLM: Conceptual understanding

To compare the results of the two-level and the single-level model, a full two-level model with all predictors of the sequential linear model (subsection 7.7.7) was calculated on the outlier-adjusted sample $(N = 365)$, see Table 116. As before, the size of significant regression coefficients and the amount of modelled variance were similar to the results

from the single-level sequential regression analysis, but p-values were slightly larger for most of the predictors.

As in the single-level model, the only significant predictors were the two novelty influence factors 'conceptual understanding in the pre-visit questionnaire' and students' physics grade with $p < 0.004$ (Bonferroni correction with $m = 12$). In total, all 12 predictors in the two-level model accounted for only approximately 32% of the variance in students' conceptual understanding in the post-visit questionnaire.

Table 116. Comparison of single-level and two-level linear regression models of novelty influence and experience factors on students' conceptual understanding: standardised regression coefficients and p-values for reduced sample without outliers ($\beta_{w/o}$, $p_{w/o}$), and the model fit criterion R^2 are displayed. Significant predictors[49] are highlighted.

	Single-level full model			two-level full model		
Predictor	$\beta_{w/o}$		$p_{w/o}$	$\beta_{w/o}$		$p_{w/o}$
(Constant)			0.305	-0.10		0.313
Age	0.03		0.556	-0.02		0.697
Concept. understanding pre UN1	**0.45**	***	4E-21	**0.44**	***	2E-19
Physics grade PH	**0.18**	**	9E-05	**0.19**	**	4E-04
Experience hands-on experiments EH	-0.08		0.052	-0.08		0.142
English skills (binary) EN	0.07		0.107	0.14		0.172
Experience with OSLEOs EO	0.02		0.614	0.01		0.827
Experience with particle physics EP	0.05		0.266	0.06		0.193
Tinkering support (novelty) TIN	0.09		0.155	0.08		0.241
Setting orientation (novelty) SO	-0.09		0.060	-0.08		0.135
Lack of problems (novelty) PRO	0.12		0.029	0.10		0.051
Cognitive load experiments CLE	-0.05		0.353	-0.04		0.478
Tutor support TS	0.03		0.543	0.05		0.456
Model fit:	$Adjusted\ R^2 = 0.39$			$Analogue\ R^2 = 0.32$		

To understand better, which of the significant predictors reduced variance at the group level and at the student level, a series of multi-level models was calculated in which the significant predictors were entered step by step. These models were then compared with an empty model without predictors and the full model with all predictors. In particular, the following four models were compared:

- **Model 0:** No predictors, only intercepts, 'empty model'
- **Model 1:** 1 predictor: pre-visit conceptual understanding
- **Model 2:** 2 predictors: pre-visit conceptual understanding & physics grade

[49] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

- **Full model for comparison:** All 12 predictors

Table 117 summarises the different models and their properties. Entering significant predictors improved the model compared to the intercept-only model 0. Specifically, every predictor improved the model compared to the previous step. However, the best model fit was reached once the two significant predictors were included. Here, model 2 accounted for 28% of the variance of the dependent variable. Adding the other 10 predictors increased the model fit parameters AIC and BIC and decreased the amount of modelled variance of the dependent variable.

In model 1, after entering students' conceptual understanding in the pre-visit questionnaire as predictor, the student-level variance decreased from 0.77 to 0.61, while, at the same time, the group-level variance decreased from 0.24 to 0.15. As expected from the results displayed in Table 101, this predictor affected variance on both levels. In model 2, the novelty influence factor 'physics grade' was entered as the second predictor, which slightly reduced the student-level variance from 0.61 to 0.59 and also slightly reduced the group-level variance from 0.15 to 0.13.

Table 117. Set of multi-level linear models on students' conceptual understanding (sample without outliers $N = 365$): Fixed part with standardised regression coefficients and p-values $(\beta_{w/o}, p_{w/o})$, and random part with level-1 and level-2 variance including standard errors (σ^2, SE). The models were compared using the model fit criteria analogue R^2, AIC and BIC.

Model	0	1		2		Full
Fixed part		$\beta_{w/o}$	$p_{w/o}$	$\beta_{w/o}$	$p_{w/o}$	**See** Table 117
Intercept	-0.03	0.00	0.984	-0.01	0.912	
Pre-visit understanding score (UN1)		**0.52**	7E-27	**0.48**	6E-23	
Physics grade (PH)				**0.22**	9E-06	
Random part						
student level σ_e^2 (SE)	0.77 (0.05)	0.61 (0.05)		0.59 (0.05)		0.57 (0.04)
group level σ_u^2 (SE)	0.24 (0.08)	0.15 (0.06)		0.13 (0.05)		0.11 (0.05)
$Analogue\ R^2$		0.24		**0.28**		0.32
AIC	1197	915		**891**		908
BIC	1205	923		**899**		915

In summary, the two novelty influence factors significantly improved this multi-level model. Together, they accounted for 28% of the variability in the dependent variable. In particular, they reduced the student-level variance from 0.77 to 0.59 and the group-level variance from 0.24 to 0.13, which is consistent with the variance components of the predictors, see Table 101.

7.8.11 Comparison of multi-level models and single-level models

In the previous section 7.8, multi-level models were presented to take the hierarchical data structure into account. In particular, students taking part in the S'Cool LAB hands-on sessions had different teachers, came in different groups, and from different countries, which was reflected in variance partition coefficients of up to 31%. Here, ignoring the hierarchical data structure translates into a dramatic overestimation of the effective sample size up to a variance inflation factor of 11 and, consequently, severely downwardly biased p-values.

By analysing variance components of the eight different educational outcomes, the clustering variable 'group' was selected as level-2 unit based on the optimal model criteria BIC and AIC. Due to the data structure (in particular, the number of level-1 units per level-2 unit) and based on theoretical considerations, random intercept models were selected to model S'Cool LAB's educational outcomes with novelty influence factors and experience factors as level-1 predictors. Here, variance partition coefficients were also calculated for all student-level (level-1) predictors to examine their potential to reduce variance at both model levels. Indeed, variance partition coefficients displayed a broad range from merely 5% (for curiosity trait) to 58% (for hands-on experience). Hence, some of the student-level predictors were expected to account for group-level variance in addition to student-level variance.

Then, full two-level models with all relevant predictors were calculated and compared with the respective single-level linear models described in section 7.6.3. As expected, adding a random intercept to the model did not significantly alter the size of the regression coefficients. Furthermore, regression coefficients of the two-levels were associated with, overall, higher p-values than regression coefficients of the single-level models, which was consistent with the smaller effective sample size in multi-level modelling. However, the conservative Bonferroni correction was already applied when interpreting significance levels within the single-level models to account for the high number of predictors as well as the deviation from a normal distribution and violations of homoscedasticity of the distribution of residuals. Therefore, almost all predictors that were significantly different from zero in the single-level models, were also significantly different from zero for the two-level models. Indeed, the p-values of most relevant predictors were very small ($p \ll 0.001$), which indicates that the chosen sample size as well as the choice and properties of the scales were well suited for these types of analyses. Furthermore, the amount of variance in the dependent variables, which was accounted for by the predictors was comparable between the two model types.

7.9 Summary: modelling S'Cool LAB's educational outcomes

In summary, the full two-level models bore a striking resemblance to the respective single-level models and confirmed the same results as summarised in subsection 7.7.8. However, p-values were underestimated in single-level models on clustered data due to the overestimation of the effective sample size. Therefore, the two-level models were more suitable for this data set than single-level models.

Theoretically, student-level predictors should only account for student-level variance when included in multi-level models. However, as expected from their variance components (see Table 101), several student-level predictors also accounted for variance at the group level. Consequently, the group-level variance was reduced significantly in all eight models. Thus, a second variance partition coefficient (VPC*) was calculated, which was based on the remaining group-level variance after including significant predictors and the total variance of the depended variable, see Table 118. Indeed, the VPC* of four of the eight educational outcomes dropped below the 5% threshold suggested by Mehmetoglu and Jakobsen (2016). Only conceptual understanding (VPC*=13%) and perceived helpfulness (VPC*=10%) demonstrated values above 10%. However, these two models also showed the lowest amount of accounted variance in general.

Table 118. Summary of two-level regression models on the eight educational outcomes. First, the estimates of the variance components of the empty models without any student-level predictors are listed including their standard error (SE) and variance partition coefficient (VPC). Next, the variance components estimates of the models that included all significant student-level predictors are listed including their standard error (SE) and variance partition coefficient (VPC*). The last row includes the sum of the accounted variance (analogue R^2). All models were based on the outlier adjusted samples.

		Interest	Curiosity	Enjoyment	Helpfulness	Self-beliefs	Easiness	Cognitive activation	Conceptual understanding
		IE3	CS3	ENJ	HEL	SC3	EAS	CA	UN3
Empty models without predictors	student level σ_e^2 (SE)	0.74 (0.05)	0.73 (0.05)	0.72 (0.05)	0.73 (0.05)	0.78 (0.05)	0.84 (0.06)	0.87 (0.06)	0.77 (0.05)
	group level σ_u^2 (SE)	0.29 (0.09)	0.19 (0.07)	0.17 (0.06)	0.24 (0.08)	0.11 (0.04)	0.15 (0.05)	0.07 (0.04)	0.24 (0.08)
	VPC	28%	21%	19%	25%	12%	15%	7%	24%
Models with only significant predictors	student level σ_e^2 (SE)	0.36 (0.02)	0.42 (0.03)	0.41 (0.03)	0.52 (0.03)	0.39 (0.03)	0.44 (0.03)	0.48 (0.03)	0.59 (0.05)
	group level σ_u^2 (SE)	0.07 (0.02)	0.04 (0.02)	0.05 (0.02)	0.10 (0.04)	0.03 (0.01)	0.02 (0.01)	0.03 (0.02)	0.13 (0.05)
	VPC*	7%	4%	6%	10%	3%	2%	3%	13%
	$Analogue\ R^2$	0.57	0.54	0.54	0.38	0.58	0.54	0.49	0.38

In summary, including suitable student-level predictors accounted for group-level differences in eight multi-level models, because the predictors themselves showed group-level variance. Depending on the variable, however, the proportion of the two variance components varied greatly. This suggests that students' reports of their subjective perception, but also of seemingly objective laboratory characteristics must be interpreted cautiously as a superposition of individual and objective components.

Last, one could argue, that multi-level modelling is not necessary if predictors account for most of the group-level variance and if the focus lies on interpreting merely significance levels. Indeed, if conservative measures are used to avoid the inflation of type I errors, for example by using the Bonferroni correction when interpreting p-values of regression coefficients, the results are very similar. However, multi-level modelling is useful to account for the reduced effective sample size that might also affect effect sizes. Furthermore, evaluating the variance-reducing behaviour of selected predictors at the different model levels provides interesting insights, as it allows to study effects within individuals as well as within groups. However, the potential of multi-level models is probably exploited better when entering both group-level and student-level predictors, or when considering more complex models, such as random slope models.

7.9.1 Summary with respect to the research questions

This section summarises the results of the single-level and two-level linear regression models to answer research questions RQ2 and RQ4.

- ***RQ2) Which of the student and laboratory characteristics are relevant for modelling the affective and cognitive educational outcomes of S'Cool LAB?***
- ***RQ4) To which extent do educational outcomes differ among students coming in different groups, from different teachers, and from different countries?***

To answer these research questions, the combined results of the single- and multi-level regression models are summarised for each of the eight educational outcomes in Table 119, and interpreted in the text below.

Here, regression coefficients are reported based on the more advanced multi-level models, which were calculated on the respective outlier-adjusted sample. The coefficients are rounded to one decimal place to account for their relatively large 95% confidence intervals that were reported together with the single-level models in section 7.7. Also, significance levels were adjusted through Bonferroni correction ($10 \leq m \leq 13$) due to the large number

of predictors that were entered into each of the models. Afterwards, predictors that were relevant for multiple educational outcomes are highlighted and interpreted based on results of similar studies. Also, intriguing differences between different regression models are discussed. Moreover, differences between students coming in different groups, from different teachers, and from different countries are highlighted.

Table 119. Summary of linear regressions on eight educational outcomes on the respective outlier-adjusted sample. Standardised significant regression coefficients are rounded to one decimal place, based on the more advanced multi-level models.

	More affective outcomes				More cognitive outcomes			
	Sit. interest	**Curiosity state**	**Enjoyment**	**Helpfulness**	**Sit. self-beliefs**	**Easiness**	**Cognitive activation**	**Conceptual understanding**
	IE3	CS3	ENJ	HEL	SC3	EAS	CA	UN3
Novelty influence factors (standardised significant regression coefficients)								
Disposition or respective 'pre' value	0.2 IE1	0.4 CS1			0.2 SC1			0.4 UN3
Other significant NIFs	0.2 CS1		0.1 IE1	0.1 IE1				0.2 PH
Novelty experience factors (standardised significant regression coefficients)								
Tinkering support (TIN)	0.3	(0.1)	0.3	0.3		0.3	0.3	
Setting orientation (SO)					0.2	0.2	0.2	
Lack of problems (PRO)						0.1		
Cognitive load experim. (CLE)					-0.3	-0.4		
Tutor support (TS)	0.3	0.2	0.3	0.3	0.2		0.3	
Multi-level model fit								
Analogue R^2	58%	54%	54%	39%	58%	53%	55%	32%
Initial group-level variance M_0	0.29	0.19	0.17	0.24	0.11	0.15	0.07	0.24
Final group-level variance M_{full}	0.07	0.04	0.05	0.10	0.03	0.02	0.03	0.13
Single-level model fit								
Adjusted R^2	55%	49%	48%	39%	52%	52%	53%	39%
Adjusted R^2 only NIFs	31%	40%	10%	11%	18%	5%	14%	36%
Adjusted R^2 only NEFs	45%	29%	47%	47%	48%	48%	46%	9%

a) Situational interest

The significant predictors among the novelty influence factors were students' dispositional interest ($\beta_{w/o} = 0.2^{***}$) and curiosity towards particle physics ($\beta_{w/o} = 0.2^{***}$), which is consistent with the theoretical background of the interest construct (Krapp, 2002) and its close relationship to epistemic curiosity (Alexander & Grossnickle, 2016). Furthermore, the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.3^{***}$), and 'tutor support' ($\beta_{w/o} = 0.3^{***}$) improved the prediction significantly. These results suggest that higher

levels of situational interest were reached by students who were already interested in physics and curious about particle physics. **Furthermore, students who perceived the learning atmosphere as favourable with high levels of tinkering and tutor support benefited more from the hands-on session in S'Cool LAB with respect to the educational outcome 'situational interest'.**

b) Curiosity towards particle physics

As expected, the strongest predictor of students' triggered curiosity was students' initial curiosity towards particle physics ($\beta_{w/o} = 0.4^{***}$). Furthermore, the novelty experience factor 'tutor support' ($\beta_{w/o} = 0.2^{***}$) improved the prediction significantly. These results suggest that higher levels of curiosity towards particle physics were reached by students who were already curious about the topic before coming to CERN and S'Cool LAB. Furthermore, students who perceived the learning atmosphere as favourable with high levels of tutor support reported higher levels of curiosity state. However, these relationships may be merely correlational: students who were already curious about particle physics beforehand might have appreciated the contact to and interaction with particle physicists more than less curious students. Also, students' curiosity towards particle physics did not change significantly from the pre- to the post-visit questionnaire.

c) Perceived enjoyment

The two strongest predictors of students' perceived enjoyment with respect to the S'Cool LAB hands-on session were the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.3^{***}$), and 'tutor support' ($\beta_{w/o} = 0.3^{***}$). Furthermore, students' dispositional interest was the third significant but weaker predictor ($\beta_{w/o} = 0.1^{*}$). These results suggest that higher levels of perceived enjoyment were reached by students who were generally interested in physics. This supports the hypothetical relationship between the two variables, with 'enjoyment' being an indicator of interest. **However, more importantly, students who perceived the learning atmosphere as favourable with high levels of tinkering and tutor support also perceived the activities in S'Cool LAB as more enjoyable.**

d) Perceived helpfulness

It comes as no surprise that the models on 'perceived helpfulness' and 'perceived enjoyment' were quite similar. After all, both variables are closely related to the interest construct. Thus, students' dispositional interest was a significant but relatively weak predictor of students' perceived helpfulness ($\beta_{w/o} = 0.1^{*}$). However, the two strongest predictors were the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.3^{***}$) and 'tutor

support' ($\beta_{w/o} = 0.3^{***}$). These results suggest that higher levels of perceived helpfulness were reached by students who were generally interested in physics. This supports the hypothetical relationship between the two variables, with 'helpfulness' serving as indicator of the 'hold' component of interest. **However, more importantly, students who perceived the learning atmosphere as favourable with high levels of tinkering and tutor support also perceived the activities in S'Cool LAB as more helpful.**

e) Situational self-beliefs

The strongest predictors of situational self-beliefs among the novelty influence factors were students' dispositional self-beliefs ($\beta_{w/o} = 0.2^{***}$), which is consistent with the theoretical background of the self-beliefs construct (Gutman & Schoon, 2013). Among the novelty experience factors, the strongest predictors were 'cognitive load of the experiments' ($\beta_{w/o} = -0.3^{***}$), 'setting orientation due to preparation' ($\beta_{w/o} = 0.2^{***}$), and 'tutor support' ($\beta_{w/o} = 0.2^{***}$). These results suggest that higher situational self-beliefs with respect to the S'Cool LAB activities were reached by students who had high dispositional physics self-beliefs. Furthermore, students who perceived the learning atmosphere in S'Cool LAB as favourable with high levels of setting orientation and tutor support, but with low levels of cognitive load related to the experiments benefited more from the hands-on workshops with respect to the educational outcome 'situational self-beliefs'. **In particular, novelty experience factors accounted for a large proportion of the variance in students' situational self-beliefs even after controlling for novelty influence factors.**

f) Perceived easiness

Although this outcome variable was correlated with the novelty influence factors 'experience with particle physics' ($r_{EP,EAS} = 0.17,\ p = 2 \cdot 10^{-4}$) and students' dispositional self-beliefs ($r_{SC1,EAS} = 0.17,\ p = 1 \cdot 10^{-4}$), only novelty experience factors accounted for significant amounts of variance of this educational outcome. In particular, the strongest predictors of the perceived easiness were 'tinkering support' ($\beta_{w/o} = 0.3^{***}$) and the perceived 'cognitive load of the experiments' ($\beta_{w/o} = -0.4^{***}$). Also, 'setting orientation due to preparation' ($\beta_{w/o} = 0.2^{***}$) and 'lack of problems' ($\beta_{w/o} = 0.1^{*}$) contributed to the regression model. In summary, these results suggest that higher levels of perceived easiness are reached by students who perceived the out-of-school learning atmosphere as favourable with high levels of tinkering support and setting orientation on the one hand, and the absence of problems as well as low levels of cognitive load on the other hand. **Furthermore, students' perception of the easiness of the S'Cool LAB hands-on session appears to be independent of novelty influence factors, such as students'**

self-beliefs or previous particle physics experiences. However, small effects of these two variables might be mediated by the novelty experience factors.

g) Perceived cognitive activation

This outcome variable was weakly correlated with the novelty influence factor 'physics grade' ($r_{PH,CA} = 0.17,\ p = 9 \cdot 10^{-5}$). Furthermore, students' perceived cognitive activation was correlated with their 'dispositional physics self-beliefs' ($r_{SC1,CA} = 0.31,\ p = 1 \cdot 10^{-12}$) and their 'dispositional physics interest' ($r_{IE1,CA} = 0.40,\ p = 2 \cdot 10^{-20}$), which supports the hypothesised relation between cognitive activation and intrinsic motivation (section 2.2.5). However, only novelty experience factors accounted for significant amounts of variance of this educational outcome. Indeed, strong predictors of students' perceived cognitive activation were the novelty experience factors 'tinkering support' ($\beta_{w/o} = 0.3^{***}$), 'setting orientation' ($\beta_{w/o} = 0.2^{***}$), and 'tutor support' ($\beta_{w/o} = 0.3^{***}$). **In summary, these results suggest that higher levels of perceived 'cognitive activation' were reached by students who perceived the out-of-school learning atmosphere as favourable with high levels of tinkering and tutor support as well as high levels of setting orientation.** Similar to the educational outcome 'perceived easiness', the perceived cognitive activation was not directly predicted by any of the novelty influence factors. Instead, the novelty experience factors seem to mediate their effects on the dependent variable.

h) Conceptual understanding

Conceptual learning posed a special case when comparing the eight regression models. Although several of the novelty influence and experience factors were weakly but significantly correlated with the outcome variable in a plausible way (e.g., English skills, experience with particle physics, tinkering support and cognitive load), only predictors modelling what students already knew accounted for variance in their post-visit conceptual understanding, which is consistent with previous results (Ausubel, 1968). In particular, strong predictors of students' conceptual understanding after taking part in the S'Cool LAB hands-on session were their initial conceptual understanding before coming to S'Cool LAB ($\beta_{w/o} = 0.4^{***}$) as well as their self-reported physics grade ($\beta_{w/o} = 0.2^{**}$). Here, the unique effect of novelty experience factors was negligible compared to these strong novelty influence factors. In particular, students' pre-visit conceptual understanding was also correlated with several novelty influence and experience factors. Therefore, it might account for the shared variance instead, as it was entered first into the regression. **In summary, these results suggest that higher levels of conceptual understanding were reached by students who already showed better understanding in the pre-visit questionnaire and who had good physics grades in school.**

Compared to the other educational outcomes, it seemed that students' perception of the learning atmosphere in S'Cool LAB was not relevant for their conceptual learning. The differences between the regression on conceptual understanding and the regressions on the other seven educational outcomes can also be interpreted by comparing the theoretical foundation of the different outcome variables. Except for conceptual understanding, all other seven outcomes merely reflect students' situational perception of themselves or of the performed activities. Conceptual understanding, on the other hand, was measured through a more objective test with multiple-choice items. Hence, although students' novelty experience was correlated with their perception of S'Cool LAB's educational outcomes, the more objective measure of their conceptual understanding might be independent of the perception of their novelty experience.

Summary across the educational outcomes

In summary, only about one third of the independent variables, which were entered into the different models, were significant predictors of the respective dependent variable. **In particular, the control variables age and gender did not account for a significant amount of variance in any of the dependent variables.** Also, students' prior experiences (with hands-on experiments, particle physics, and out-of-school learning) and English skills were not relevant for any of the linear regressions on the educational outcomes. Although some of these novelty influence factors were weakly but significantly correlated with some of the outcomes, these novelty influence factors did not manifest themselves as significant predictors. Probably, their small effects were accounted by other predictors, which shared variance with both the respective novelty influence factor and the respective outcomes variable. **Overall, the results suggest that educational outcomes of S'Cool LAB were independent of students' English skills and prior experiences.** However, the remaining novelty influence factors, namely dispositions or other corresponding variables in the pre-visit questionnaire, accounted for significant amounts of variance in their situational counterparts. Indeed, these novelty influence factors accounted for about one third of the variance in the outcomes 'situational interest', 'curiosity state towards particle physics', and 'conceptual understanding'.

Overall, novelty experience factors accounted with about 40%, for twice the amount of variance in the dependent variables than did novelty influence factors. This suggests that the characteristics of the out-of-school learning laboratory were more important than student characteristics with respect to students' educational outcomes. Only for two outcomes, namely 'curiosity state' and 'conceptual understanding', did novelty influence factors account for a higher amount of variance than novelty experience factors. Here, students' curiosity towards particle physics was already very high in the pre-visit

questionnaire and did not increase further after participation in the hands-on session in S'Cool LAB. Thus, it is plausible that the pre-visit curiosity state strongly predicts the post-visit curiosity state. When considering students' conceptual understanding, it is well know that the strongest predictor of learning is what students already know (Ausubel, 1968; Hattie, 2008). Moreover, in both of these two cases, the items assessing the outcomes in the post-visit questionnaire were already included in a very similar way in the pre-visit questionnaire, which also explains the high amount of shared variance between these two pairs of variables. In general, several novelty influence factors correlated with novelty experience factors, and their respective modelled variance showed an overlap. Therefore, novelty experience factors seem to mediate effects of several novelty influence factors as suggested by Cors (2016).

Among the novelty experience factors, the support variables 'tinkering support' (e.g., helpful instructions and easy to find materials) and 'tutor support' (e.g., help in case of problems and display of fascination for physics) were the most relevant predictors: both were significantly different from zero in six of the eight models. In particular, these variables demonstrated high regression coefficients for the more affective outcomes, namely 'situational interest', 'curiosity towards particle physics', perceived 'enjoyment' and 'helpfulness' as well as for the outcome 'cognitive activation'. **Hence, students' perception of support – both by the learning environment and the educators in the room – were crucial elements of the novelty experience S'Cool LAB, particularly with respect to the affective effects of the hands-on sessions in S'Cool LAB and students' perceived cognitive activation.** This is consistent with findings by Pawek (2009), who identified support by tutors and a positive learning atmosphere as promising predictors for situational interest. Also, Glowinski and Bayrhuber (2011) identified the support by scientists ('quality of instruction') as important predictor of situational interest. Moreover, Simon (2019) confirmed the importance of support variables for students' interest. Here, the other affective outcomes and cognitive activation are closely related to the construct 'interest', which can explain similarities in the respective regression models.

The interpretation of the predictor 'lack of problems' was difficult, as it was based only on three (inversely coded) items and it suffered from a suppression effect for the regression on perceived helpfulness, see 7.7.3. In any case, this predictor demonstrated only comparatively small regression coefficients.

Setting orientation due to students' preparation (e.g., students knew what to expect in S'Cool LAB and were familiar with the schedule) was relevant for the more cognitive outcomes. However, it did not contribute significantly to modelling the affective outcomes.

This suggests that preparation had no negative (or positive) effects on students' interest, enjoyment, or curiosity. Instead, it was being beneficial for several cognitive outcomes, namely 'situational self-beliefs', perceived 'easiness' and 'cognitive activation'. Similarly, students' perceived cognitive load (e.g., students found the experiments difficult and the instructions hard to understand) was a significant predictor of the more cognitive outcomes 'situational self-beliefs' and perceived 'easiness'. **Hence, students' perception of their (cognitive) preparedness and their cognitive load were vital components of the novelty experience in S'Cool LAB, in particular with respect to its cognitive effects.** This is consistent with previous findings, for example by Streller (2015), who confirmed positive effects on students' situational self-beliefs (and interest) when preparing them via an online portal. Also, Molz (2016) confirmed that preparation aiming at reducing setting novelty had positive effects on motivation, which included a self-beliefs component.

Several authors suspected that the high amount of novelty in out-of-school learning labs affects learning and development (Cors et al., 2017; Guderian, 2007; Itzek-Greulich, 2015). However, only few studies took students' perceptions of their novelty experience into account or considered factors that might influence novelty experience. **Based on the results of the S'Cool LAB main study, it can be summarised that the novelty influence and experience factor model described by Cors (2016) is a very useful tool when modelling educational outcomes of an out-of-school learning opportunity: it provides a theoretical framework to include students' and laboratory characteristics in a systematic way.** Furthermore, both types of novelty factors and their different dimensions account for high amounts of variance in several educational outcomes. In particular, the S'Cool LAB main study showed that novelty factors affected educational outcomes most if they belonged to similar dimensions. For example, novelty factors belonging to the cognitive dimension and factors of the dimension 'setting familiarity' were mostly relevant for cognitive outcomes, while novelty factors belonging to the affective dimension, as well as the social and 'capabilities' dimension were mostly relevant for affective outcomes.

However, these considerations only applied to self-reported educational outcomes, such as students' perceived cognitive activation. The objectively measured outcome 'conceptual understanding' seemed to be independent of students' novelty experience. This suggests that the learning effect of S'Cool LAB hands-on sessions is indeed more robust than the other educational outcomes. However, the amount of variance of the outcome variable 'conceptual understanding' that was accounted for by predictors was comparatively low. It is likely that this model, in particular, would have benefited from additional novelty influence or experience factors.

Compared to the related 'interest' construct, the regression models on 'self-beliefs' stand out for several reasons. First, the variance partition coefficient was much smaller for dispositional physics self-beliefs than for dispositional physics interest (15% vs. 28%), which indicates that group variables such as the teacher have a smaller effect on students' dispositional self-beliefs compared to their interest. Second, as the only significant novelty influence factor, dispositional physics self-beliefs only accounted for 18% of the variance in students' situational self-beliefs, which is much lower than the 31% of variance in their situational interest accounted for by novelty influence factors. This could indicate that students perceived the learning environment in S'Cool LAB differently from their physics classes. Indeed, Osborne et al. (2003) described "*the apparent contradiction between students' attitudes towards science in general and their attitudes towards school science*". Here, students' perceptions of themselves in physics classes might be different than their perception of themselves in physics activities outside school. Third, novelty experience factors alone accounted for almost all the variance in students' situational self-beliefs. This suggests that a favourable novelty experience at an out-of-school learning lab might lead directly to high situational self-beliefs, almost independent of students' dispositional self-beliefs. Considering the impact of self-beliefs on course choices and career aspirations (Nagy et al., 2008; Parker et al., 2014), these results make the variable 'self-beliefs' particularly interesting for out-of-school learning opportunities.

Students taking part in S'Cool LAB hands-on sessions had different teachers, came in different groups, and from different countries, which was reflected in variance partition coefficients of up to 31% with respect to the educational outcomes. Here, ignoring the hierarchical data structure translated into a dramatic overestimation of the effective sample size up to a variance inflation factor of 11 and, consequently, severely downwardly biased p-values. When comparing the three different clustering variables, however, the clustering according to 'group' led to the best model fit criteria for seven of the eight educational outcomes. Furthermore, even the predictors that were considered student-level variables, reflected the hierarchical data structure with variance partition coefficients between merely 5% (for curiosity trait) and 58% (for hands-on experience). In summary, students who came to S'Cool LAB on the same day as part of the same group shared similar novelty influence factors, for example their previous experience with hands-on experiments or their dispositional physics interest. Adding these novelty influence factors as student-level predictors into two-level models accounted for most of the group differences with respect to the educational outcomes. In contrast, novelty experience factors differed far less between different groups. This suggests that students from different groups perceived their novelty experience in S'Cool LAB in a similar way. Furthermore, group differences in

students' educational outcomes can mostly be accounted for by group differences in students' novelty influence factors.

Overall, in particular affective educational outcomes and students' conceptual understanding differed significantly among students coming in different groups, from different teachers, and from different countries, because the student groups differed significantly in their dispositional interest and their prior conceptual understanding.

8 Study on students' conceptions of radiation & radioactivity

As summarised in subsection 7.4.4, students' taking part in hands-on sessions in S'Cool LAB significantly improved their conceptual understanding of selected key concepts of the three workshops. For this purpose, seven multi-tier multiple-choice items were adapted based on previous concept test items or constructed based on documented students' conceptions and evaluated within the S'Cool LAB pilot study (see section 6.4). Although the properties of these items were satisfactory, their information content was limited. In particular, these items focus on only three separate key concepts and due to the very restricted testing time available the item number was limited. Also, the lack of existing research and challenges presented by a specific vocabulary that is mostly unfamiliar to high-school students interfered with the desire to measure students' conceptual understanding of particle physics concepts consistently.

Here, the topic of radiation and radioactivity provides an interesting alternative because it is based on the conceptual understanding of elementary particles and their fundamental interaction with other particles. Furthermore, students have prior experience with this topic from formal and informal learning opportunities. However, while research about students' conceptions of radiation and radioactivity started already in 1975, no suitable test instrument exists to assess these conceptions in a quantitative way. Moreover, most of the previous studies in this field are comparatively old, and it is not obvious, which of the previously documented conceptions are still relevant for today's learners. Therefore, a different set of students in S'Cool LAB was used to expand our current understanding of students' conceptions of ionising radiation and radioactivity and to develop and validate a suitable quantitative test instrument.

For this, a multiple-true-false item format was chosen based on considerations, which are described in the next section (8.1). The development of these MTF items is presented in section 8.2. Next, section 8.3 introduces the sample used for this item development and the handling of missing data. Furthermore, section 8.4 summarises the item analysis. Moreover, items were grouped into different categories according to the two-level source-radiation-absorber model, see section 8.5. Subsequently, section 8.6 presents the results of an exploratory factor analysis that shows which items grouped together empirically and compares the empirical results with the theoretical categorisation. The results of the factor analysis then formed the basis of a cluster analysis presented in section 8.7, which examined whether students can be grouped into clusters according to their answer behaviour with respect to the identified factors. Finally, section 8.8 summarises the findings with respect to the associated research questions.

8.1 Why Multiple-True-False items?

Many concept inventories rely on Multiple-Choice (MC) items with only one correct answer option and several distractors, which are based on previously identified students' conceptions. However, MC items have a number of disadvantages compared to other item formats, such as Multiple-True-False (MTF). For example, designing suitable distractors for MC items requires a lot of effort. For some concepts, it might not even be possible to provide the desired number of effective distractors, which may have negative effects on the test's internal consistency (Kaplan & Saccuzzo, 2009). Nevertheless, even when students succeed in identifying the correct answer option, they might not completely discard the other options if they had the chance to. In fact, students often demonstrate an incoherent understanding of scientific concepts, which DiSessa (1988) describes as an unstructured collection of phenomenological primitives (p-prims). MTF items have the potential to reveal this partial understanding by identifying even a mixture of accurate and inaccurate conceptions (Brassil & Couch, 2019; Parker et al., 2012).

MTF items typically include an item stem followed by a number of statements, and participants need to decide for every statement, whether it is true or false. Therefore, this item format is better-suited and more accurate in measuring complete mastery of a topic, as well as in revealing students' alternative conceptions (Brassil & Couch, 2019; Couch, Hubbard, & Brassil, 2018). Furthermore, MTF items are easy to construct, administer and score (Kaplan & Saccuzzo, 2009). As for disadvantages, Kaplan and Saccuzzo (2009) list a higher number of items required to compensate for the high guessing ratio as well as forcing students into a 'black and white thinking' that may not be suitable for certain concepts. With respect to guessing, Cronbach (1941) suggests that *"students should be told, before taking a test, the percentage of items keyed "true", so that they may adjust their responses accordingly"*.

Based on a) the number of loosely connected students' conceptions related to radiation and radioactivity (see subsection 2.4.2), b) the problems associated with constructing suitable distractors for some concepts (see section 6.4) and c) the findings from the studies listed above, this chapter describes the development and evaluation of a set of MTF items to assess students' conceptions of radiation and radioactivity.

8.2 Item development

A set of 103 MTF items was developed based on the documented students' conceptions about radiation and radioactivity, which are described in more detail subsection 2.4.2. Whenever possible, previously tested multiple-choice items were adapted or extended. In particular, the developed items focus on radioactive decay processes, characteristics of radioactive atoms, half-life, ionisation, types of ionising radiation, and the interaction of radiation with matter. Table 120 lists the 12 different item stems as well as the sources of the respective items. The final test booklet including the item stems and the full item texts is included in appendix 10.10.

Table 120. List of the 12 different item stems together with the sources of the respective items.

Item stem		Item source
A	There are different types of radioactive decays, every type is characterised by a certain process in the atom.	Based on Prather (2000)
B	What characterises radioactive atoms?	Based on Johnson (2014), Neumann and Hopf (2013); Prather (2000)
F	What is the half-life?	Based on Eijkelhof (1990); Prather (2000)
G	A substance (mass 100 g, volume 80 cm3) is made of 1000 radioactive atoms with a half-life of 1 hour. How many radioactive atoms are left after 1 hour? How many atoms in total are left after 1 hour?	Based on Prather (2000)
H	Which properties of a substance determine the half-life of the radioactive atoms in this substance?	Based on Tekin and Nakiboglu (2006)
I	What is ionisation?	Based on Colclough (2007); Johnson (2014)
J	There are different types of ionising radiation.	Based on (Johnson, 2014); Neumann and Hopf (2012)
K	Can the following types of radiation ionise an atom?	Based on Eijkelhof (1990)
L	How can you turn stable atoms into radioactive atoms?	Based on Eijkelhof (1990); Tekin and Nakiboglu (2006)
M	Doctors forgot to switch off the X-ray machine in a hospital. Inside the X-ray machine, they forgot a ring made of copper. What happens to the copper when is it exposed to X-rays for a 1 hour?	Based on Colclough (2007); Eijkelhof (1990); Riesch and Westphal (1975)
N	Doctors forgot to switch off the X-ray machine in a hospital. Inside the X-ray machine, they forgot a patient. What happens to the body of the patient when is it exposed to X-rays for a 1 hour?	Based on Colclough (2007); Eijkelhof (1990); Henriksen and Jorde (2001); Millar (1993); Riesch and Westphal (1975)
O	How dangerous radiation is for humans depends on …	Based on Johnson and Maidl (2014)

8.3 Sample and handling of missing data

Due to an already high test load caused by items assessing students' various educational outcomes and perceptions of novelty factors within the framework of hands-on sessions in S'Cool LAB, no additional items could be included in the main study presented in chapter 7. Instead, the developed multiple-true-false items were evaluated with a separate student sample. In particular, data collection took place in the framework of S'Cool LAB's new research programme, the so-called S'Cool LAB PLUS+ programme, between October 2018 and June 2019 with 363 students. Due to the external constraints of this research programme, the time available for having students fill out paper questionnaires was limited to 40 minutes directly before starting the hands-on workshops in S'Cool LAB. Hence, some of the students did not finish the questionnaires in time. Since factor analysis is susceptible to missing values, careful analysing and handling of missing values is therefore required beforehand. Table 121 summarises this process.

In particular, 61 of the 363 students (17%) missed answering more than 4 of the items (>4%). Typically, their missing data was clustered at the last part of the questionnaire, which was an indicator for lack of time to fill out the test caused by, for example, low English comprehension skills, slow reading, or slow processing. Consequently, these students were excluded from further analysis, which might have biased the sample slightly towards high achievers and limits the comparability of the items' easiness scores. Also, 67 of the 363 students (18%) missed answering only a few items (1 to 4 items, <4%) and showed a random pattern of their missing items. This indicates that these students forgot to answer or did not know how to answer a specific item. Hence, these missing answers were coded as wrong (0) and included in further analysis. In this way, data from 302 students (average age $17.2\ y, SD = 0.7y$, 35% female) was used for further analysis.

Table 121. Procedure for handling missing data. In total, 363 students filled out the questionnaire with 93 items.

# Missing items	# Students (% of 363)	Procedure
1	45 (12%)	Coded as wrong (0), 102 substitutions (67 students, 18%)
2	13 (4%)	
3	5 (1.4%)	
4	4 (1.1%)	
5	3 (0.8%)	Excluded from further analysis (61 students, 17%)
6	3 (0.8%)	
7	2 (0.6%)	
8	3 (0.8%)	
>8	50 (14%)	
For analysis:	**302 (83%)**	

These 302 students came in 19 groups from 11 different countries (Finland 8%, France 4%, Germany 23%, Israel 1%, Italy 10%, Luxembourg 8%, Norway 6%, New Zealand 4%, Poland 6%, Portugal 11%, United Kingdom 20%). The selection of student groups for S'Cool LAB's research programme S'Cool LAB PLUS+ focused on student and teacher motivation and on an English level of at least B1. Therefore, all students should have been able to read and understand the questions presented. Also, the author was available to answer questions about item wording while students filled out the questionnaires. However, as discussed above, students with lower English reading comprehension skills might have needed more time to fill out the questionnaire, and some of them have been excluded from further analysis if they were missing more than four items. Due to the age demographic of the S'Cool LAB PLUS+ programme, most groups consisted of advanced level physics classes, which explains the gender bias with only 35% female students.

8.4 Item analysis

Although all items were based on documented conceptions and many items were adapted from previous multi-choice items, all items were analysed to evaluate the test's face validity. Here, the item analysis was based on interviews with four students and a discussion with five physics education experts, which helped to identify items that were ambiguous or did not assess students' thinking as intended. Consequently, problematic items were excluded from further analysis. Furthermore, the easiness of all items was assessed within the sample described above.

8.4.1 Interviews and expert discussion

In October 2018, directly after filling out the questionnaires, four students were randomly selected to take part in brief follow up interviews. In these approximately 20-minute-long one-on-one interviews, the MTF items were discussed one by one together with the interviewees to identify items with difficult or problematic wording as well as items without a clear correct answer. Furthermore, all item texts were discussed with five physics education experts at CERN in August 2019. Consequently, 11 of the 102 items were discarded, see Table 163 in appendix 10.11 for the detailed list of excluded items.

For example, item A8 ("*During one type of radioactive decay, the atom breaks up into 2 stable atoms of approximately the same size*.") led students to think about spontaneous nuclear fission, which was not intended to be part of the questionnaire. Item I5 ("*During ionisation, ions emit radiation*."), item J6 ("*Some types of ionising radiation are emitted by ions*.") as well as item J7 ("*Some types of ionising radiation can transform stable atoms*

into radioactive atoms.") were excluded because they are ambiguous and can be true or false depending on the exact circumstances. Item K1 ("*Can the following types of radiation ionise an atom: Heat?*") was dropped for the same reasons; depending on how thermal radiation is defined, it can comprise ionising components, for example, by including thermal radiation of stars. Furthermore, item L7 was adapted after the analysis of the interviews. Here, students' answer behaviour directly reflects the changes in the item wording:

- **Version 1:** "*Stable atoms will become radioactive if you bring them in direct contact with a radioactive source for a while*." False, easiness: $M = 32\%, n = 170$
- **Version 2:** "*Stable atoms can become radioactive if you put them in a plastic bag and place the bag in direct contact with a radioactive source for a while*." False, easiness: $M = 48\%, n = 132$

Indeed, 68% of the students thought that direct contact to a radioactive source will transform stable atoms into radioactive atoms. After adding a hypothetical sheet of plastic between the source and the stable atoms in the item text, only 52% thought that the stable atoms will become radioactive. This suggests that students may have confused irradiation and contamination in this case. However, due to these changes to the item text, this item was excluded from further analysis.

Two more items were ambiguous for experts upon closer inspection but were not excluded from the analysis because it was very unlikely that students would have the knowledge required to detect their ambiguity. In particular, the answer to item L6 ("*Stable atoms can become radioactive if you place them in a 10 cm distance to a radioactive source for a while*.") is only ambiguous if students consider neutron radiation (for example by alpha-beryllium neutron sources). Item G1 ("*After 1 hour (= 1 half-life), the total number of atoms is 1500, because half of the radioactive atoms decayed into 2 stable atoms*.") could be answered as "TRUE", for example, when considering a pure alpha-emitter such as Polonium-210 transforming into the stable isotope lead-206 and a stable helium nucleus.

8.4.2 Item easiness

After excluding items based on the interviews with students and discussion with experts, the remaining 92 items were analysed with respect to their easiness for the study sample. Table 122 displays a short version of the item texts, the source of the items as well as their easiness within the study sample described above. The original test booklet including the full item texts and item stems is included in appendix 10.10. In addition, one column lists

the item category that was assigned through an expert rating. Details about these categories and the categorisation process are included in section 8.5.

In total, ten items (items B3, F2, J8, K13, K16, L4, M9, N7, O2, O5) were very easy with more than 90% correct answers. For example, only 5% of the students expected a patient to 'glow in the dark' after being X-rayed for an hour (N7) and only 9% expected copper to 'glow in the dark' after one hour in an X-ray machine (M9). Also, merely 3% of the students answered that the half-life is the time in which a radioactive source is dangerous (F2). Moreover, only 9% thought that heat could turn stable atoms into radioactive atoms (L4). Finally, only 8% of the students answered that an imbalance between the number of electron and neutrons causes atoms to be radioactive (B3). Previous studies documented all these conceptions, but it seems that these conceptions were not prevalent among the students, who took part in this study.

On the other hand, eight items were very difficult with easiness scores below the guessing ratio of 50% (items A2, A3, A5, L2, L5, L6, N5, and O6). For example, only 45% of the students answered correctly that during one type of radioactive decay, an electron is emitted from the nucleus (A2). Furthermore, 48% answered correctly that an electron drops to a lower energy level during one type of radioactive decay (A3). Only 28% of the students answered correctly that the nucleus drops to a lower energy level during one type of radioactive decay (A5). Also, 49% of the students answered correctly that X-rays (L5) or a radioactive source in 10 cm distance (L6) cannot turn stable atoms into radioactive atoms. One of the most challenging items was item N5. Here, only 38% of the students answered correctly that X-radiation does not accumulate inside the body if a patient is exposed to X-radiation for an hour. These results suggest strong misconceptions among the students, who took part in this study, which are consistent with previous studies.

Table 122. List of 92 MTF items including label, whether they were true T or false F, short description, source and easiness. The column 'level' includes the assigned item category, a combination of Source (S), Radiation (R), Absorber (A) as well as the levels sub-microscopic (mi) or macroscopic (ma). Items with easiness ≥ 90% are highlighted. See appendix 10.10 for item details.

label	T/F	Short description	Level	Source	Easiness
A1	F	Decay = an electron is emitted from the orbital	S-mi	Based on Prather (2000)	56%
A2	T	Decay = an electron is emitted from the nucleus	S-mi		45%
A3	F	Decay = an electron drops to lower energy level	S-mi		48%
A4	T	Decay = p^+ and n^0 are emitted from the nucleus	S-mi		68%
A5	T	Decay = the nucleus drops to lower energy level	S-mi		28%
A6	F	Decay = the atom slowly disappears	S-mi		85%
A7	F	Decay = the nucleus slowly disappears	S-mi		83%
B1	F	Radioactive = unbalanced number of e^- and p^+	S-mi	Based on Johnson (2014), Neumann and Hopf (2013); Prather (2000)	73%
B2	T	Radioactive = unbalanced number of n^0 and p^+	S-mi		69%
B3	F	Radioactive = unbalanced number of e^- and n^0	S-mi		92%
B4	T	Radioactive = excited state with too much energy	S-mi		67%
B5	F	Radioactive = atoms already existed for too long	S-mi		87%
B6	F	Radioactive = atoms in contact with chemicals	S-mi		67%
B7	F	Everything is a little bit radioactive	S-mi		61%
F1	F	T1/2 = the nucleus loses half of its radiation	S-mi	Based on Eijkelhof (1990); Prather (2000)	70%
F2	F	T1/2 = time in which a substance is dangerous	S-ma		97%
F3	F	T1/2 = radioact. substance loses half its mass	S-ma		55%
F4	F	T1/2 = time in which the nucleus splits into 2 parts	S-mi		89%
F5	T	T1/2 = half the radioactive atoms decayed	S-ma		79%
F6	F	T1/2 = radioact. substance loses half its volume	S-ma		70%
G1	F	T1/2 = half the atoms decayed into 2 stable atoms	S-ma	Based on Prather (2000)	83%
G2	F	T1/2 = 40cm^3 left of 80 cm^3 radioact. substance	S-ma		72%
G3	F	T1/2 = 50g left of 100g radioact. substance	S-ma		60%
G5	T	T1/2 = 500 of 1000 radioact. atoms transformed	S-ma		68%
G6	F	T1/2 = same nr. of atoms but weaker radiation	S-ma		79%
G7	F	T1/2 = 500 of 1000 atoms disappeared	S-ma		80%
H1	F	T1/2 depends on the temperature of a substance	S-ma	Based on Tekin and Nakiboglu (2006)	78%
H2	F	T1/2 depends on the pressure of a substance	S-ma		80%
H3	F	T1/2 depends on the density of a substance	S-ma		66%
H4	F	T1/2 depends on previous contact to chemicals	S-ma		77%
H5	T	T1/2 depends on the properties of the nucleus	S-mi		85%
H6	F	T1/2 depends on the age of a substance	S-ma		79%
H7	F	T1/2 depends on the properties of the orbital	S-mi		62%
I1	T	Ionisation = atoms loose or gain electrons	A-mi	Based on Colclough (2007); Johnson (2014)	88%
I2	F	Ionisation = stable atoms become radioactive	A-mi		84%
I3	F	Ionisation = e^- excited to higher energy level	A-mi		85%
I4	F	Ionisation = nucleus excited to higher energy level	A-mi		86%
I6	F	Ionisation = atoms loose or gain protons	A-mi		84%
J1	T	Radiation can be described by particle model	R-mi	Based on (Johnson, 2014); Neumann and Hopf (2012)	80%
J2	T	Radiation can be described by wave model	R-mi		70%
J3	F	Radiation can consist of radioactive particles	R-mi		57%
J4	F	Radiation can consist of pure energy	R-mi		77%
J5	T	Radiation can be electrically charged	R-mi		63%
J8	T	Radiation can be dangerous for humans	R-ma		90%

label	T/F	Short description	Level	Source	Easiness
K2	F	Radio waves … can ionise	R-ma	Based on Eijkelhof (1990)	81%
K3	F	Mobile phone radiation … can ionise			88%
K4	T	Gamma radiation … can ionise			87%
K5	T	Cosmic radiation … can ionise			71%
K6	T	Beta minus radiation (electrons) … can ionise			80%
K7	T	Alpha radiation (Helium nuclei) … can ionise			71%
K8	T	X-rays … can ionise			76%
K9	F	Microwaves … can ionise			68%
K10	F	Laser light from a red laser pointer … can ionise			83%
K11	F	Infrared radiation (IR) … can ionise			70%
K12	T	Ultraviolet Light (UV light) … can ionise			63%
K13	F	Visible light … can ionise			90%
K14	F	Ultrasound used in hospitals … can ionise			80%
K15	F	Radiation from batteries … can ionise			76%
K16	F	Wi-Fi radiation …can ionise			94%
L1	F	Contact to chemicals can make atoms radioactive	A-mi	Based on Eijkelhof (1990); Tekin and Nakiboglu (2006)	64%
L2	T	Neutron exposure can make atoms radioactive	A-mi		40%
L3	T	Proton exposure can make atoms radioactive	A-mi		58%
L4	F	Heat can make atoms radioactive	A-mi		91%
L5	F	X-radiation can make atoms radioactive	A-mi		49%
L6	F	Source (10cm away) can make atoms radioactive	A-mi		49%
M1	F	Copper gets warm when X-rayed 1h	A-ma	Based on Colclough (2007); Eijkelhof (1990); Riesch and Westphal (1975)	66%
M2	F	Copper smells when X-rayed 1h	A-ma		87%
M4	F	X-radiation accumulates in front of copper	A-ma		78%
M5	F	X-radiation accumulates inside of copper	A-ma		54%
M6	F	X-radiation is reflected by the copper surface	A-ma		82%
M7	T	Copper electrons are excited when X-rayed 1h	A-mi		75%
M8	T	Chemical bonds break when X-rayed 1h	A-mi		53%
M9	F	Copper glows in the dark when X-rayed 1h	A-ma		91%
M10	F	Copper becomes radioactive when X-rayed 1h	A-mi/ma		60%
N1	F	Body gets warm when X-rayed 1h	A-ma	Based on Colclough (2007); Eijkelhof (1990); Henriksen and Jorde (2001); Millar (1993); Riesch and Westphal (1975)	60%
N2	F	Body smells when X-rayed 1h	A-ma		83%
N4	F	X-radiation accumulates in front of the body	A-ma		84%
N5	F	X-radiation accumulates inside the body	A-ma		38%
N6	F	X-radiation is reflected by bones in the body	A-ma		80%
N7	F	Patient glows in the dark when X-rayed 1h	A-ma		95%
N8	T	DNA is destroyed in some cells when X-rayed 1h	A-mi		85%
N9	F	Person gets poisoned when X-rayed 1h	A-ma		67%
N10	T	Some cells will die when X-rayed 1h	A-ma		89%
N11	F	Person will lose hair when X-rayed 1h	A-ma		73%
N12	F	Person becomes radioactive when X-rayed 1h	A-mi/ma		72%
N13	T	Some cells will start to mutate when X-rayed 1h	A-ma		77%
O1	T	Radiation danger depends on energy	A-ma	Based on Johnson and Maidl (2014)	77%
O2	T	Danger of radiation depends on the radiation type	A-ma		92%
O3	T	Danger of radiation depends on the intensity	A-ma		88%
O4	T	Danger of radiation depends on the distance	A-ma		88%
O5	T	Danger of radiation depends on exposure time	A-ma		94%
O6	T	Danger of radiation depends on the body part	A-ma		44%

8.5 Classification of item categories

The items listed above were developed to represent the three aspects 'Source' (S), 'Ionising radiation' (R) and 'Absorber' (A) listed in Figure 38 on both the sub-microscopic level (label: mi) and the macroscopic level (label: ma). This Figure was already presented and explained in section 2.4.1. In particular, items were expected to group according to the following six categories:

- Source of ionising radiation: S-mi and S-ma
- Ionising radiation: R-mi and R-ma
- Absorber of ionising radiation: A-mi and A-ma

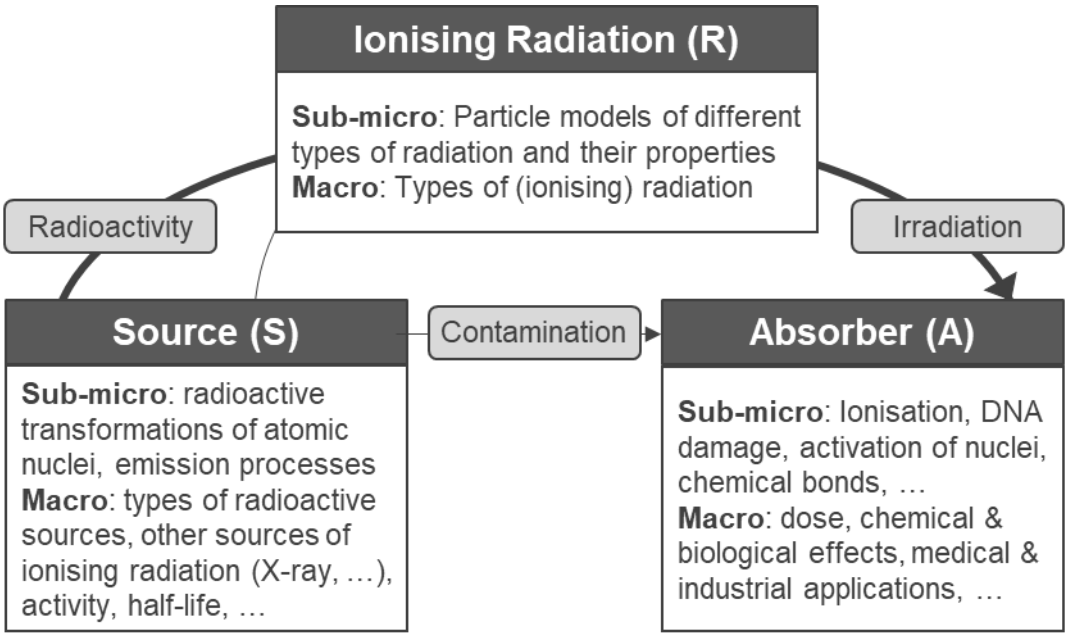


Figure 38. Extended framework of radiation & radioactivity concepts at macroscopic and sub-microscopic levels based on (Millar et al., 1990).

To evaluate the construct validity of the test, physics education experts at CERN were asked to assign the remaining 92 items listed in Table 122 to one of these six categories. In particular, during a discussion with five physics education experts at CERN, a coding frame was developed. This coding frame used Figure 38 together with the following three rules:

1. Item categories are assigned based on what learners will most likely think about when answering each item.
2. The sub-microscopic level includes single molecules (including DNA), single atoms, as well as chemical bonds between pairs of molecules/atoms, and everything smaller (atomic nucleus, orbital, electron, …). Everything bigger than single molecules is considered to belong to the macroscopic level.

3. Item set K2 to K16, which includes a list of different types of (ionising) radiation is treated as one item during the categorisation process.

Following these rules, statements about the half-life, such as F1 "*The half-life is the time in which the nucleus of a radioactive atom loses half of its radiation*." were assigned to category S-mi (source, sub-microscopic level). Although half-life is considered a macroscopic concept relating to an ensemble of radioactive atoms, this item mentions "*the nucleus*" and is therefore expected to focus students' thinking on the sub-microscopic level.

One additional physics education expert rated all 92 items based on Figure 38 and the above listed coding rules to calculate the interrater agreement as well as Cohen's kappa coefficient (κ) (Landis & Koch, 1977) between this physics education expert and the author. The Initial rating showed a substantial agreement of 82% (Cohen's $\kappa = 0.77$). After a discussion, two additional coding rules were added for the following cases:

4. If an item addresses ionising radiation together with an absorber (or a source), the item is coded from the viewpoint of the absorber (or the source).
5. If an item includes a statement about absorbers becoming radioactive, both codes A-mi and A-ma will be used, as students' reasoning depends on the conception of radiation they hold. Although activation is considered a sub-microscopic process, many students hold rather macroscopic views of radiation and how it accumulates inside an object.

After these adjustments, both raters showed an almost perfect agreement of 92% (Cohen's $\kappa = 0.90$). However, disagreement persisted for the following six items:

- B7 ("*All atoms are radioactive because everything is a little bit radioactive*."): Rater 2 assigned category S-ma instead of S-mi, stating that the word "*everything*" can be perceived as a strong trigger word for the macroscopic world.
- G5, G6, G7 (Number of atoms left after 1 half-life when starting with 1000 radioactive atoms): Rater 2 assigned category S-mi instead of S-ma, stating that students will most likely think about single atoms even if the large atom numbers are used as trigger words for the macroscopic level.
- M6 & N6 (X-radiation is reflected by copper surface / by bones in the body): Rater 2 assigned category A-mi instead of A-ma, stating that students will think of reflection as a sub-microscopic process.

In summary, physics education experts agreed when assigning the 92 MTF items to one of the six categories of the two-level source-radiation-absorber model.

8.6 Exploratory factor analysis

Although a relatively long list of loosely connected students' conceptions identified by previous studies is included in subsection 2.4.2, our knowledge of students' reasoning patterns is limited. Which of the items presented above correlate? Can groups of correlated items be interpreted according to uniting themes? Do certain conceptions of sub-microscopic processes correlate with certain conceptions of macroscopic effects or do items group according to the hypothetical categories presented in section 8.5? To answer these questions, an exploratory factor analysis was conducted on the binary data generated by the MTF items.

However, since factor analysis is sensitive to the ratio of item number and number of cases, several of the 92 MTF items listed in Table 122 were excluded to obtain a more stable factor solution. In particular, 10 items that were very easy (easiness $\geq 90\%$) were excluded because they would provide little interesting information and little variance to analyse. Also, the remaining 13 items of item stem K ("*Can the following types of radiation ionize an atom? Radio waves, Mobile phone radiation, Gamma radiation*, …") were excluded from factor analysis because these items provide information rather about students' knowledge than about their conceptual understanding. Instead, these items were analysed separately. However, difficult items (easiness $< 50\%$) were included in the factor analysis because they provide interesting information about students' misconceptions while at the same time offering enough variance for factor analysis.

Consequently, the covariance of 69 items was analysed through principal axis factoring with a sample size of 302 students resulting in a case to item ratio of 4. Osborne et al. (2008) analysed 303 case-to-item ratios in exploratory factor analyses and found that 41% of the studies report a ratio $\leq$ 5:1. Here, a case to item ratio of 4:1 for this study is not critically low. Nevertheless, factor analysis would benefit from a larger sample size. Therefore, all of the extracted factors must be interpreted cautiously.

Because of the dichotomous data, the correlation matrix was produced using the tetrachoric item correlation in STATA 13. First, an exploratory factor analysis (principal axis factoring) was computed to estimate the number of factors by inspecting the scree plot of the resulting eigenvalues, see Figure 39. The scree plot shows three main components whose eigenvectors lie on a line with the same slope. However, the first three factors accounted only for 21% of the variance in the 69 items. In total, 31 factors with eigenvalues greater than 1 were identified. When assuming that the item group according to the hypothetical categories presented in section 8.5, the 69 items should form six factors.

Furthermore, Figure 39 shows a small kink after factor ten, which could be an indicator of the factor number. Here, the first ten factors account for 50% of the variance in the items, see Table 123.

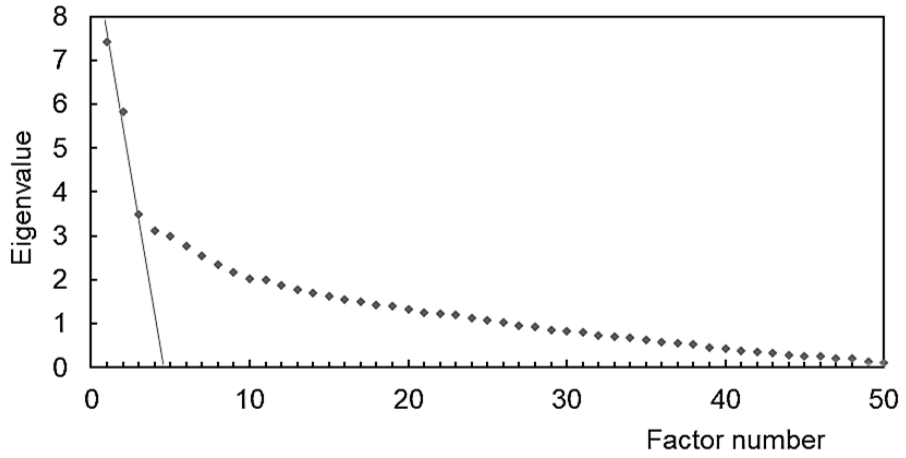


Figure 39. Scree plot of eigenvalues of exploratory principal axis factoring.

Therefore, several factor solutions for three to ten factors were computed and compared to identify the most suitable ('cleanest') factor structure following the procedure suggested by Osborne et al. (2008). In particular, factor solutions were rotated orthogonally through Varimax rotation with Kaiser Normalisation to ease the interpretation of the factors. Very low factor loadings $|a| < 0.32$ were omitted, and items without relevant factor loadings or with several high cross-loadings were not interpreted unless there was strong theoretical support for a specific factor assignment.

Table 123. Results of exploratory principal axis factoring showing the first ten of the extracted factors including their eigenvalues, % of the modelled variance, and cumulative modelled variance.

Factor	Eigenvalue	% modelled variance	cumulative modelled variance
1	7.4	10.6%	10.6%
2	5.8	8.3%	18.9%
3	3.5	5.0%	23.9%
4	3.1	4.5%	28.4%
5	3.0	4.3%	32.6%
6	2.8	4.0%	36.6%
7	2.6	3.6%	40.3%
8	2.3	3.3%	43.6%
9	2.2	3.1%	46.7%
10	2.0	2.9%	49.6%

Furthermore, items with low item communalities ($h^2 < 0.4$) were discarded from a factor if they did not seem to fit because low item communalities indicated that these items were either not related to the other items or that they might require an additional factor.

Following this procedure, the seven-factor solution presented below fitted the data of the 69 items best. In this solution, six of these seven factors can be interpreted in a meaningful way. However, 11 Items did not fit any of the following seven factors.

1. Basic knowledge
2. Sub-microscopic basis of the radioactive decay
3. Half-life = half the source (S-ma)
4. Irradiation vs. contamination
5. Chemical properties of sources and invisible absorber effects
6. Tangible macroscopic radiation effects (A-ma)
7. (Potpourri of sub-microscopic misconceptions)

In the following, the extracted factors are presented in more detail.

8.6.1 Factor 1: Basic knowledge

Factor 1 included 16 items. These 16 items were categorised into five different categories, see Table 124, ranging from the sub-microscopic level description of radioactive decay (S-mi) to the dangers of ionising radiation for macroscopic human absorbers (A-ma).

Interestingly, all of the 16 statements were correct and had to be rated ‘true’. In other words, none of these 16 items referred to a misconception. Therefore, the items seemed to form a factor consisting of basic knowledge of radioactivity and radiation. Not all of the students in the sample had studied the topic in depths within their physics classes yet. Therefore, this factor could maybe also be interpreted as ‘school knowledge of radioactivity and radiation’. Indeed, students who learned about, for example, the half-life in school, might have also learned about different types of radiation and the dangers of radiation, and vice versa.

Several of the items showed cross-loadings on other factors. However, these cross-loadings were typically smaller than the loadings for factor 1 or negative. In this case, negative factor loadings could indicate another sub-factor within this factor. Item I1 showed similar loadings for factors 1 and 7 but was assigned to factor 1 for two reasons: this item fits the interpretation of the factor as basic knowledge and factor 7 could not be interpreted in a meaningful way.

Table 124. Factor 1: Basic knowledge. This table lists item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings $|a| < 0.32$ are omitted.

	TF	Item description	Cat	f1	f2	f3	f4	f5	f6	f7	h^2
A4	T	Decay = p^+ and n^0 are emitted from the nucleus	S-mi	0.52							0.3
B2	T	Radioactive = unbalanced number of n^0 and p^+	S-mi	0.56	0.39						0.5
F5	T	T1/2 = half the radioactive atoms decayed	S-ma	0.35						-0.46	0.4
G5	T	T1/2 = 500 of 1000 radioactive atoms transformed	S-ma	0.36						-0.38	0.4
H5	T	T1/2 depends on the properties of the nucleus	S-mi	0.51	0.32		-0.40	0.37			0.7
I1	T	Ionisation = atoms loose or gain electrons	A-mi	0.37						0.41	0.4
J1	T	Radiation can be described by particle model	R-mi	0.50							0.4
J2	T	Radiation can be described by wave model	R-mi	0.55							0.5
J5	T	Radiation can be electrically charged	R-mi	0.56							0.4
L2	T	Neutron exposure can make atoms radioactive	A-mi	0.58	0.39						0.5
M7	T	Copper electrons are excited when X-rayed 1h	A-mi	0.31					-0.33		0.3
N8	T	DNA is destroyed in some cells when X-rayed 1h	A-mi	0.34							0.3
O1	T	Radiation danger depends on energy	A-ma	0.58							0.4
O3	T	Radiation danger depends on the intensity	A-ma	0.59							0.4
O4	T	Radiation danger depends on the distance	A-ma	0.52							0.4
O6	T	Radiation danger depends on exposed body part	A-ma	0.55							0.4

8.6.2 Factor 2: Sub-microscopic basis of the radioactive decay (S-mi)

Factor 2 included 10 items, eight of which had to be rated 'false', see Table 125. Eight of the 10 items referred to the sub-microscopic level of radiation sources. They covered, for example, reasons for and the process of radioactive decay. In particular, three items addressed the concept of 'half-life'. Although half-life is a macroscopic physics concept, two of the statements included trigger words such as 'nucleus' and 'orbital', which aimed to focus students' thinking towards the structure of atoms and their role in radioactive transformation processes. Only item H4 (half-life depends on the previous contact to chemicals) came as a surprise in this factor. Although it showed high factor loading for factor 2, it was previously assigned to category S-ma, the macroscopic level of radiation sources. However, students might correctly connect chemical reactions with the orbital of atoms and apply reasoning on the sub-microscopic level instead. The same could apply to item L1 (contact to chemicals can make atoms radioactive), which was expected to belong to the absorber category A-mi. Since this statement referred to absorber atoms

transforming into radioactive atoms (and thereby into a radioactive source), it makes sense to find this item within this factor. Furthermore, both of the items H4 and L1 that referred to the effect of chemicals on radioactive sources showed additional loadings on factor 5, which consisted of several items referring to the chemical properties of radioactive sources. However, the loadings on factor 1 were higher for these two items, which was used to assign them to factor 2. In addition, item H7 showed a high additional loading on factor 6. However, it fitted factor 2 better from a theoretical point of view, since all other items in factor 6 referred to macroscopic effects of radiation on absorbers.

Table 125. Factor 2: Sub-microscopic basis of the radioactive decay (S-mi). This table lists item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings $|a| < 0.32$ are omitted.

	T F	Item description	Cat.	f1	f2	f3	f4	f5	f6	f7	h^2
A1	F	Decay = an electron is emitted from the orbital	S-mi		0.72						0.5
A2	T	Decay = an electron is emitted from the nucleus	S-mi		0.70						0.6
A3	F	Decay = an electron drops to lower energy level	S-mi		0.54						0.3
A5	T	Decay = the nucleus drops to lower energy level	S-mi		0.43						0.3
B1	F	Radioactive = unbalanced number of e^- and p^+	S-mi		0.51						0.4
B6	F	Radioactive = atoms were in contact with chemicals	S-mi		0.47						0.3
F4	F	T1/2 = time in which the nucleus splits into 2 parts	S-mi		0.54						0.3
H4	F	T1/2 depends on previous contact to chemicals	S-ma		0.52			0.39			0.5
H7	F	T1/2 on the properties of the orbital	S-mi		0.43				0.38		0.5
L1	F	Contact to chemicals can make atoms radioactive	A-mi		0.52			0.34			0.5

8.6.3 Factor 3: Half-life = half the source (S-ma)

Factor 3 included six items, all of which had to be rated ‘false’, see Table 126. This factor can be interpreted as a ‘half-life’ factor and includes mostly items from category S-ma that refer to students’ conceptions of half-life. In particular, it entails students’ conceptions of half-life as the time in which half of the atoms, and hence also half of the mass and volume of a substance disappear. Interestingly, item A6 (Decay = the atom slowly disappears, category S-mi) shows the highest loading on this factor with a high cross-loading on factor 2. Indeed, it would be consistent if students hold a rather continuous than discrete conception of the radioactive decay, and also associate ‘decay’ with the disappearance of a substance, see Jansky (2019). Therefore, it was decided to assign this item to factor 3, instead of to factor 2.

Item L5 (Atoms can become radioactive during exposure to X-rays) loads highly on this factor. However, it does not seem to belong to the other items from a theoretical point of view and might be an artefact of the factor analysis and the relatively small sample size. Therefore, it was excluded from this factor and not interpreted within any other factor.

Table 126. Factor 3: Half-life = half the source (S-ma). This table lists item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings $|a| < 0.32$ are omitted.

	T F	Item description	Cat.	f1	f2	f3	f4	f5	f6	f7	h^2
A6	F	Decay = the atom slowly disappears	S-mi		0.44	0.47					0.5
F3	F	T1/2 = radioactive substance loses half its mass	S-ma			0.80					0.7
F6	F	T1/2 = radioactive substance loses half its volume	S-ma			0.75					0.6
G2	F	T1/2 = 40cm^3 left of 80 cm^3 radioactive substance	S-ma			0.76					0.6
G3	F	T1/2 = 50g left of 100g radioactive substance	S-ma			0.68					0.5
G7	F	T1/2 = 500 of 1000 atoms disappeared	S-ma			0.49					0.4
~~L5~~	~~F~~	~~X-radiation can make atoms radioactive~~	A-mi			0.38					0.4

8.6.4 Factor 4: Irradiation vs. contamination

Factor 4 consisted of six items, all of which had to be rated 'false' to be solved, see Table 127. Five of the six items were previously assigned to the absorber categories (A-mi and/or A-ma). This factor can be summarised as 'irradiation versus contamination'. Although students' conceptions related to this area are well-known and documented in previous studies, it is promising to analyse the clustering of items in this factor. Items M5 and N5 refer to the 'accumulation of radiation' inside the absorber. These items follow the often-held conception of matter-like properties of radiation. Here, some students might think that the absorber will 'soak up radiation like a sponge' and release it later, thereby acting as a radiation source itself. Items L6, M10 and N12 refer to the absorber turning into a source of ionising radiation without mentioning a specific process, such as 'accumulation'. Item J3 (Radiation can consist of radioactive particles) does not fit this factor well when considering its assigned category R-mi and its quite low communality ($h^2 = 0.3$). However, it would be consistent if students, who think of radiation as the transfer of 'radioactive particles', also think that exposure to 'radioactive particles' will turn the absorber into a radiation source. Indeed, this reasoning is absolutely correct when considering contamination instead of an irradiation process. However, more (qualitative) research would be needed to clarify whether this factor resembles consistent student reasoning on the sub-microscopic level based on the 'radioactive particles' conception of radiation.

Table 127. Factor 4: Irradiation vs. contamination. This table lists item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings $|a| < 0.32$ are omitted.

	T F	Item description	Cat.	f1	f2	f3	f4	f5	f6	f7	h^2
J3	F	Radiation can consist of radioactive particles	R-mi				0.39				0.3
L6	F	Source (10cm away) can make atoms radioactive	A-mi				0.38				0.4
M5	F	X-radiation accumulates inside the copper	A-ma				0.64				0.5
M10	F	Copper becomes radioactive when X-rayed 1h	A-mi/ A-ma				0.57				0.5
N5	F	X-radiation accumulates inside the body	A-ma				0.67				0.5
N12	F	Person becomes radioactive when X-rayed 1h	A-mi/ A-ma				0.53				0.4

8.6.5 Factor 5: Chemical properties of sources and invisible absorber effects

Factor 5 consisted of 10 items, eight of which had to be rated 'false' to be solved, see Table 128. The items originate from three different hypothesized categories (S-mi, S-ma, and A-ma). Here, items H1, H2, H3 and H6 refer to external factors or chemical properties, such as temperature, pressure, density and the age of a substance, which affect the half-life (category S-ma). Items M4, N4, N6, N10 and N13 refer to macroscopic absorbers, either describing radiation accumulating in front of an absorber (M4, N4), radiation being reflected by an absorber (N6), or describing actual biological effects on cells in a human absorber (N10, N13). Together, these items can be summarised within two groups: chemical properties of sources and invisible absorber effects. However, it is difficult to find a link between these two groups. Furthermore, item A7 is difficult to interpret within this factor. It might be merely an artefact, as it refers to the sub-microscopic conception (S-mi) of radioactive decay involving the nucleus to disappear slowly. However, expecting a nucleus to disappear slowly during radioactive decay might be the basis to believe that chemical properties are relevant for this process: maybe some students think of decay as being similar to chemical dissolution processes? Here, the temperature would indeed affect the decay time. In summary, more (qualitative) research is needed to clarify whether this item is an artefact or fits into a coherent reasoning pattern.

Items N4 (radiation accumulates in front of the body) shows similar high loading on factor 4. However, since items N4 and M4 were constructed in a very similar way, it was decided to keep them together within factor 5. Other cross-loadings were either negative (M4, N10) or much smaller (N13) and thus tolerated.

Table 128. Factor 5: Chemical properties of sources and invisible absorber effects. This table lists item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings $|a| < 0.32$ are omitted.

	TF	Item description	Cat.	f1	f2	f3	f4	f5	f6	f7	h^2
A7	F	Decay = the nucleus slowly disappears	S-mi					0.42			0.3
H1	F	T1/2 depends on the temperature of a substance	S-ma					0.52			0.4
H2	F	T1/2 depends on the pressure of a substance	S-ma					0.54			0.4
H3	F	T1/2 depends on the density of a substance	S-ma					0.44			0.3
H6	F	T1/2 depends on the age of a substance	S-ma					0.37			0.3
M4	F	X-radiation accumulates in front of copper	A-ma					0.37	-0.35		0.4
N4	F	X-radiation accumulates in front of the body	A-ma				0.46	0.40			0.5
N6	F	X-radiation is reflected by the bones in the body	A-ma					0.45			0.2
N10	T	Some cells will die when X-rayed for 1 hour	A-ma					0.35	-0.34		0.4
N13	T	Some cells will start to mutate when X-rayed for 1 hour	A-ma	0.33				0.45			0.4

8.6.6 Factor 6: Tangible macroscopic radiation effects (A-ma)

Factor 6 consisted of six items, all of which had to be rated 'false' to be solved, see Table 129. All six items were previously assigned to the macroscopic absorber category (A-ma). They include statements about macroscopic radiation effects on absorbers that can be perceived with human senses (smell, warmth, poisoned, hair loss). Thus, they can be summarised as 'tangible macroscopic radiation effects'.

Table 129. Factor 6: Tangible macroscopic radiation effects (A-ma). This table lists item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings $|a| < 0.32$ are omitted.

	TF	Item description	Cat.	f1	f2	f3	f4	f5	f6	f7	h^2
M1	F	Copper gets warm when X-rayed 1h	A-ma						0.49		0.3
M2	F	Copper smells when X-rayed 1h	A-ma						0.62		0.5
N1	F	Body gets warm when X-rayed 1h	A-ma						0.62		0.5
N2	F	Body smells when X-rayed 1h	A-ma						0.68		0.6
N9	F	Person gets poisoned when X-rayed 1h	A-ma						0.50		0.3
N11	F	Person will lose hair when X-rayed 1h	A-ma						0.39		0.2

8.6.7 Factor 7: Potpourri of sub-microscopic misconceptions

Factor 7 consisted of six items all of which had to be rated 'false' to be solved, see Table 130. Factor 7 includes items about the source (B7, F1, G1, G8), the absorber (I3) and radiation (J4), mostly referring to the sub-microscopic level. Only items G1 and G6 were initially assigned to a macroscopic level (S-ma) by the author. However, the second rater suggested category S-mi for these two items. Two of the items show negative loadings (F1, G6) on this factor, which indicates that they might form a sub-factor. Hence, they were not interpreted within this factor.

Table 130. Factor 7: Potpourri of alternative sub-microscopic conceptions. This table lists item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings $|a| < 0.32$ are omitted. Items F1 and G6 (struck through) are not interpreted here due to negative loadings.

	T F	Item description	Cat.	f1	f2	f3	f4	f5	f6	f7	h^2
B7	F	Everything is a little bit radioactive	S-mi							0.32	0.2
G1	F	T1/2 = half the atoms decayed into 2 stable atoms	S-ma							0.32	0.2
I3	F	Ionisation = electrons excited to higher energy level	A-mi		0.34					0.53	0.5
J4	F	Radiation can consist of pure energy	R-mi							0.56	0.4
~~F1~~	~~F~~	~~T1/2 = the nucleus loses half of its radiation~~	~~S-mi~~							~~-0.44~~	~~0.3~~
~~G6~~	~~F~~	~~T1/2 = same number of atoms but weaker radiation~~	~~S-ma~~							~~-0.67~~	~~0.7~~

There seems to be no theme uniting these items, which is not surprising considering that this is the last of the extracted factors. Also, this factor only accounts for 4% of the item variance and two of the items show very low loadings and very low communalities (B7, G1). Furthermore, item I3 shows a high cross-loading on factor 2. Hence, these items were summarised as 'potpourri of sub-microscopic conceptions' and excluded from further analysis.

8.6.8 Items that were not interpreted within the seven-factor solution

In addition to the problematic factor 7, 11 of the 69 items could not be interpreted within the 7-factor solution, see Table 171 in appendix 10.12.8. Eight of these items demonstrated communalities below 0.4. Furthermore, six items did not load on any of the factors (B4, B5, I2, I6, M6, M8). In contrast, item I4 showed high loadings on four different factors and, thus, was not interpreted. Items L3, F1 and G6 showed only negative loadings on one of the extracted factors. Item L5 was excluded from factor 3 because it did not seem to fit and demonstrated low item communality.

8.6.9 Correlations between the extracted factors

Orthogonal factor rotation techniques, such as Varimax assume uncorrelated factors. However, analysing the correlations between the extracted factors can still provide first information about students' reasoning patterns. Therefore, Table 131 shows the correlation matrix of factors 1 to 6. Factor 7 was excluded because there was no uniting theme for the items in this factor.

Indeed, four pairs of factors were correlated. Here, students' conceptions related to the 'sub-microscopic basis of the radioactive decay' (factor 2) were weakly correlated with three other factors, namely with students' basic knowledge (factor 1, $r = 0.28$***), with students' conceptions of half-life (factor 3, $r = 0.30$***), and with students' conceptions related to chemical properties of sources and invisible absorber effects (factor 5, $r = 0.20$*). Furthermore, factor 3 (half-life) and factor 4 (irradiation vs. contamination) were weakly correlated ($r = 0.23$*). Indeed, it would be consistent if students with a good understanding of sub-microscopic processes of radioactive sources would also understand other radiation concepts better. However, all four correlations were small and previous research suggests that students demonstrate a rather incoherent understanding of scientific concepts (DiSessa, 1988), with a mixture of accurate and inaccurate conceptions (Brassil & Couch, 2019; Parker et al., 2012). Nevertheless, the results of the factor analysis suggest that students argue somewhat consistently within the small topic areas represented by the 6 extracted factors. Consequently, a cluster analysis was calculated to determine whether different groups exhibit different reasoning patterns with respect to the 6 extracted factors, see section 8.7.

Table 131. Pearson correlation coefficients between extracted factors. Significance levels adjusted via Bonferroni correction ($m = 21$): * $p < 0.002$, ** $p < 0.0004$, *** $p < 0.00004$.

Factors	1	2	3	4	5	6
1 Basic knowledge	1	**0.28*** **	-0.05	0.01	0.17	-0.15
2 Sub-microscopic basis of the radioactive decay (S-mi)		1	0.11	**0.30*** **	**0.20***	0.08
3 Half-life = half the source (S-ma)			1	**0.23***	0.13	0.09
4 Irradiation vs. contamination				1	0.15	0.13
5 Chemical properties of sources and invisible absorber effects					1	0.08
6 Tangible macroscopic radiation effects (A-ma)						1

8.6.10 Summary of factor analysis and scale properties

The previous sections described the results of an exploratory factor analysis of 69 MFT items about radiation and radioactivity. In summary, the 7-factor solution was favoured as it displayed the most suitable and 'cleanest' factor structure. Consequently, six of the seven factors were interpreted in a meaningful way based on a common underlying physics concept or theme. In addition, factor 5 appeared to consist of two sub-factors, namely the chemical properties of radioactive sources and invisible absorber effects.

The clustering of the items is also reflected in the internal consistency of the factors, see Table 132. Indeed, the first three factors show good internal consistency with Cronbach's $\alpha_C \geq 0.7$. On the other hand, factors 4, 5 and 6 demonstrate relatively poor internal consistency with Cronbach's α_C only slightly above 0.5. However, these scales consisted of individual MTF items from different item stems with somewhat low communalities. Therefore, a low internal consistency comes as no surprise and, nevertheless, provides a promising basis for further scale development. Table 132 provides an overview of the scale properties of the six interpreted factors as well as of the seventh extracted factor that was excluded before. Furthermore, appendix 10.12 lists the corrected item-to-total correlation of all items as well as Cronbach's alpha when single items were omitted from the respective scale.

Table 132. Overview of the scale properties of the seven extracted factors including the factor description, the number of items per factor (#), the scale's mean (M) and standard deviation (SD), Cronbach's α_C including the 95% confidence interval and the amount of variance accounted for.

	Description	#	M	SD	α_C	95% CI	%variance
1	Basic knowledge of radiation and radioactivity	16	0.73	0.18	0.71	[0.66, 0.76]	8%
2	Sub-microscopic basis of the radioactive decay (S-mi)	10	0.61	0.24	0.70	[0.65, 0.75]	7%
3	Half-life (S-ma)	6	0.70	0.28	0.70	[0.64, 0.75]	6%
4	Irradiation vs. contamination	6	0.55	0.27	0.57	[0.48, 0.64]	5%
5	Chemical properties of sources and invisible absorber effects	10	0.79	0.17	0.51	[0.42, 0.59]	5%
6	Tangible macroscopic radiation effects (A-ma)	6	0.73	0.24	0.55	[0.47, 0.62]	5%
7	(Potpourri of sub-microscopic students' conceptions)	4	0.77	0.24	0.33	[0.20, 0.45]	4%
-	Not interpreted	11	-		-		-

It is difficult to compare these results to the results obtained by previous studies because factor analyses were rarely conducted in this area. However, Molz (2016) identified a similar 'irradiation vs. contamination' factor when analysing his achievement test. Colclough (2007) did not conduct a factor analysis but reports the results of an expert

rating. Indeed, three experts were asked to assign 21 MC items to the three quite similar categories 'absorption/penetration', 'irradiation/ contamination', and 'micro-scale models related to radioactivity and ionising radiation'. Finally, Boyes and Stanisstreet (1994) reported a factor analysis based on 1365 students who answered 36 five-level items about the origin, movement, uses, and dangers of radioactivity and radiation. They interpreted the following eight factors after an exploratory factor analysis: 'telecommunications and transmission', 'common perception of dangers', 'gaseous pollution', 'therapeutic and hygienic uses', 'rays"', 'science fiction', 'penetrating capacity' and 'nuclear power station emissions'. However, their test items showed barely any similarity to the items within this study and can thus not be compared.

8.6.11 Types of radiation types: ionising versus non-ionising

Items K2 to K16 ("*Can the following types of radiation ionize an atom?*") were not included in the exploratory principal axis factoring. Instead, they were analysed separately, see Table 133.

Table 133. Can the following types of radiation ionise an atom? Students answered 'yes' or 'no' to 15 types of radiation, some of them were fictional (ultrasound, radiation from batteries).

	Can the following types of radiation ionise an atom?	Correct answers	% of students answering 'Yes'	Easiness
K14	Ultrasound used in hospitals	No	20%	80%
K15	Radiation from batteries	No	24%	76%
K2	Radio waves	No	19%	81%
K3	Mobile phone radiation	No	12%	88%
K16	Wi-Fi radiation	No	6%	94%
K9	Microwaves	No	32%	68%
K11	Infrared radiation (IR)	No	30%	70%
K10	Laser light from a red laser pointer	No	17%	83%
K13	Visible light	No	10%	90%
K12	Ultraviolet Light (UV light)	Yes	63%	63%
K6	Beta minus radiation (electrons)	Yes	80%	80%
K8	X-rays	Yes	76%	76%
K4	Gamma radiation	Yes	87%	87%
K7	Alpha radiation (Helium nuclei)	Yes	71%	71%
K5	Cosmic radiation	Yes	71%	71%

On average, students answered 79% of these items correctly ($SD = 0.14$). It was most difficult for them to recognise microwaves as non-ionising radiation (69% correct answers) and to recognise UV light as ionising radiation (63% correct answers). In contrast, students easily identified Wi-Fi (94% correct answers) and visible light (90% correct answers) as types of non-ionising radiation. Among the types of ionising radiation, gamma radiation was identified correctly by most of the students (87%), which is consistent with previous

results suggesting that students perceive gamma radiation as more dangerous than other types of ionising radiation, such as X-rays (Mubeen et al., 2008). Alarmingly, almost a quarter of the students (24%) expected batteries to emit ionising radiation. Also, 30% said that infrared radiation would be able to ionise an atom. Although only 6% of the students expected Wi-Fi radiation to ionise, 12% answered that mobile phone radiation can ionise an atom. In summary, the majority of the students was able to distinguish ionising from non-ionising radiation types. However, even the small proportion of students who struggled to distinguish these radiation types is worrying, as these students will not be able to gauge associated dangers (or the absence of dangers) appropriately.

8.7 k-means cluster analysis

When analysing students' answers to questionnaire items, factor analysis helps to identify clusters of items that were answered in a similar way, which could indicate a common underlying concept or theme. Cluster analysis, on the other hand, can help to identify clusters of students who answered in a similar way. In the following, the results of a k-means cluster analysis on students' mean values for factors 1 to 6 are presented. As discussed before, factor 7 was excluded because of its weak scale properties and weak theoretical background. The k-means algorithm maximises the variance between a given number of sub-groups, see subsection 5.3.2. Here, cluster analysis was performed five times to obtain 2 to 6 clusters. After comparing these five solutions, the 3-cluster solution seemed most promising and, thus, is summarised in Table 134 and presented below.

Table 134. Results of *k*-means clustering for 3 clusters. Convergence was reached after 15 iterations. The final cluster centres m_1, m_2 and m_3 are compared to the overall mean value of each factor (M) and are highlighted as follows: green $m > 0.7$, yellow $0.5 \leq m \leq 0.7$, red $m < 0.5$.

#	Factor	Factor mean M	Cluster 1 (n = 118) m_1	2 (n = 75) m_2	3 (n = 109) m_3
1	Basic knowledge	0.73	0.79	0.75	0.66
2	Sub-microscopic basis of the radioactive decay (S-mi)	0.61	0.79	0.51	0.48
3	Half-life = half the source (S-ma)	0.70	0.79	0.35	0.86
4	Irradiation vs. contamination	0.55	0.77	0.33	0.46
5	Chemical properties of sources and invisible absorber effects	0.79	0.86	0.74	0.76
6	Tangible macroscopic radiation effects (A-ma)	0.73	0.79	0.68	0.69

Furthermore, Figure 40 shows groups of boxplots of the mean values across the six different factors and three student clusters.

In this solution, **cluster 1** consists of 118 students, who performed well on most of the items, which resulted in factor mean values well above average. **Cluster 2** consists of 75 students, who demonstrated average basic knowledge of radiation and radioactivity (factor 1) but a poor understanding of the sub-microscopic basis of the radioactive decay (factor 2) and slightly below-average mean values of factor 5 (chemical properties of sources and invisible absorber effects) and 6 (tangible macroscopic radiation effects). However, students in cluster 2 demonstrated severe misconceptions with respect to the half-life (factor 3) and irradiation vs. contamination (factor 4). Finally, **cluster 3** consists of 109 students with lower basic knowledge (factor 1), who performed well on the items forming factor 3 (half-life), factor 5 (chemical properties of sources and invisible absorber effects) and factor 6 (tangible macroscopic radiation effects). However, students in cluster 3 demonstrated a poor understanding of factor 2 (sub-microscopic basis of the radiative decay) and factor 4 (irradiation vs. contamination).

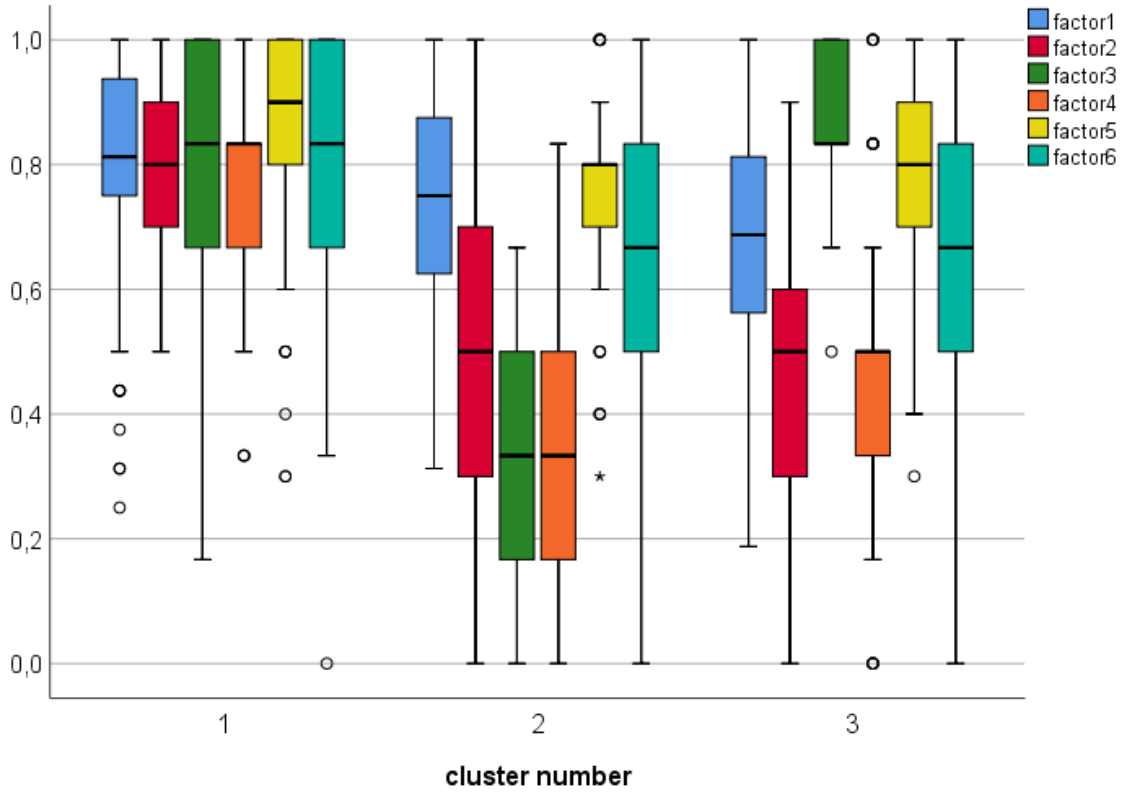


Figure 40. Boxplots of the mean values across three student clusters and the six different factors: factor 1: Basic knowledge, factor 2: Sub-microscopic basis of the radioactive decay (S-mi), factor 3: Half-life = half the source (S-ma), factor 4: Irradiation vs. contamination, factor 5: Chemical properties of sources and invisible absorber effects, and factor 6: Tangible macroscopic radiation effects (A-ma).

8.8 Summary with respect to the research questions

A set of multi-true-false items was developed based on documented students' conceptions as a first attempt to measure and analyse students' conceptual understanding of radiation and radioactivity systematically and quantitatively. Here, factor analysis identified interesting clusters of items, most of which could be interpreted in a meaningful way as internally consistent scales. While correlations among the extracted factors were very weak, a cluster analysis identified thought-provoking reasoning patterns. In particular, severe misconceptions about certain areas were independent of each other and independent of students' basic knowledge. In the following, key results are interpreted with respect to the relevant research questions introduced in section 3.3.

► ***RQ5.1) Which of the students' conceptions identified by previous studies are relevant to today's learners taking part in S'Cool LAB workshops?***

Only very few of the items generated based on previously documented students' conceptions were very easy for students taking part in S'Cool LAB workshops and, consequently, were answered correctly by at least 90% of the students. For example, very few students expected a patient or a copper object to 'glow in the dark' after being X-rayed for an hour. Indeed, this conception was identified by Boyes and Stanisstreet (1994), but, 26 years later, this conception seems to be no longer prevalent. Also, almost none of the students believed that the half-life is the time in which a radioactive source is dangerous, or that heat can turn stable atoms into radioactive atoms. Moreover, only a small fraction of the students answered that an imbalance of the number of electrons and neutrons causes atoms to be radioactive. Furthermore, very few students said that Wi-Fi radiation or visible light could ionise atoms.

However, the majority of the items was sufficiently difficult for students taking part in S'Cool LAB workshops. In particular, eight items were very difficult, with success rates below the guessing rate of 50%. In particular, students struggled to identify the role of electrons in radioactive transformations and to distinguish phenomena of the nucleus from phenomena of the orbital. Instead of expecting an electron being emitted from the nucleus, more than half of the students answered that an electron drops to a lower energy level during one type of radioactive decay. Also, almost half of the students expected that an electron would be emitted from the orbital during radioactive decay. Furthermore, students confused irradiation and contamination, which might be based on a wrong model of radiation as consisting of radioactive particles. Consistently with this wrong model, an overwhelming number of students answered incorrectly that X-radiation accumulates inside the body if a patient is exposed to X-radiation for an hour.

These results suggest that students' conceptual understanding of radiation and radioactivity has not changed much over the past decades. Instead, most of the previously documented conceptions, which formed the basis of the MTF items, were reproduced within this study. Furthermore, many misconceptions seem to be based at the sub-microscopic level.

► ***RQ5.2) What are the similarities and differences between a hypothetical two-level source-radiation-absorber model and the factor structure of the developed Multiple-True-False items?***

Based on the source-radiation-absorber model, documented students' conceptions were categorised into six categories, which consisted of the three concepts 'source', 'radiation' and 'absorber' each on a sub-microscopic level and a macroscopic level. Consequently, the MTF items used in this study were developed to represent selected aspects for each of the six categories. Here, it was assumed that factor analysis would reproduce this hypothetical structure and confirm that students argue consistently within one concept and level, but not necessarily across different concepts and levels. Consequently, learners might struggle to link the two different levels similar to the level confusion described with respect to the 'chemistry triplet' (Johnstone, 1991). This level confusion might then even account for common misconceptions.

Here, the results of the factor analysis only partially reproduced the expected structure. Three of the extracted factors can clearly be assigned to one of the six categories. In particular, factor 2 represents the sub-microscopic level of radioactive sources including the reason for and the process of decay processes. Also, factor 4 consists of items referring to the macroscopic concept of half-life, which is associated with radioactive sources. Last, factor 6 refers to macroscopic effects radiation may have on absorbers.

However, the factors seem to represent rather small areas of consistent student thinking than distinguishable physics concepts. For example, three items related to the macroscopic concept of half-life were identified to be part of the sub-microscopic factor 2. These three items included trigger words that might focus students thinking about half-life on the sub-microscopic level. Consequently, they were more similar to the other items of the sub-microscopic factor 2 than to the macroscopic factor 3 (half-life). Furthermore, one item that referred to the sub-microscopic conception of a rather continuous than discrete decay process was associated with factor 3, which otherwise includes only items referring to the macroscopic level of the concept 'half-life'.

Three of the extracted factors could not be assigned to one of the six categories. Instead, they consisted of items referring to different concepts and levels. In particular, factor 5 seems to consist of two sub-factors referring to the macroscopic level of radioactive sources and absorbers of radiation. Moreover, factor 4 refers to irradiation and contamination, which link the concepts source, radiation and absorber. Here, factor 1 stands out as it entails basic knowledge of radiation and radioactivity across all categories and levels. However, this factor consists of almost all items that had to be rated 'true' because they referred to true statements rather than misconceptions. To some degree, general knowledge and specific groups of misconceptions were separated through factor analysis.

► ***RQ5.3) What are common reasoning patterns exhibited by clusters of students?***
In summary, students taking part in this study seemed to demonstrate neither an unstructured collection of phenomenological primitives as proposed by DiSessa (1988) nor a coherent understanding of scientific concepts. Instead, the results of the factor analysis suggest that, on the one hand, students argue rather consistently within certain areas, such as the macroscopic perspective of half-life. On the other hand, the results of the cluster analysis suggest that students' conceptual understanding of one area is independent of their understanding of a different area. In particular, students in two of the three clusters described above exhibited severe misconceptions with respect to two factors, while, at the same time, showing average or above-average understanding with respect to other factors. Most strikingly, these misconceptions seem to manifest themselves independently of the students' knowledge. Here, MTF items proved their potential to reveal this partial understanding by identifying even a mixture of accurate and inaccurate conceptions.

9 Discussion and conclusions

S'Cool LAB provides a unique research setting, and its development as well as the evaluation of its educational impact on students are relevant for educators involved both in out-of-school learning opportunities as well as in formal education. The following summary (9.1) comprises a reflection on the main results of the S'Cool LAB studies and their implications for out-of-school learning as well as formal education in general. Then, section 9.2 discusses strengths and limitations of the presented S'Cool LAB studies. Last, section 9.3 highlights aspects of the S'Cool LAB studies that could serve as starting points for future research.

9.1 Summary and implications

9.1.1 Educational outcomes

In summary, the S'Cool LAB studies show that the topic of particle physics provides an excellent opportunity to introduce high-school students to modern physics while taking their interest and curiosity into account without compromising conceptual understanding. Indeed, the high educational effectiveness more than justifies the required resources.

Medium-sized effect on physics interest

The results of the S'Cool LAB studies confirm high educational effectiveness of its hands-on sessions. Even though the students, who took part in the hands-on sessions in S'Cool LAB were already interested in physics beforehand (66% of the maximum possible score), they demonstrated still higher levels of situational interest with respect to the activities in S'Cool LAB (77% of the maximum possible score). Indeed, despite the short intervention time of only 4.5 hours, participation in a half-day hands-on session in S'Cool LAB led to a medium-sized effect on students' physics interest ($d_C = 0.6$). This results is consistent with previous results, for example, by Molz (2016) and Pawek (2009). However, the absolute values are higher than previously reported values, both for the dispositional and situational interest. Indeed, in a study with more than 11 000 students, Priemer et al. (2018) found an average value of 59% for the epistemic component of situational interest after taking part in hands-on out-of-school learning labs. In summary, S'Cool LAB performs very well compared to other hands-on out-of-school learning labs by triggering very high situational interest even for students whose dispositional interest was already very high.

Effect on physics self-beliefs

Similarly, students' physics self-beliefs were already relatively high before coming to S'Cool LAB (66% of the maximum possible score) but students reported even higher levels of situational self-beliefs after their hands-on session in S'Cool LAB (74% of the maximum possible score). Here, participation in a hands-on session in S'Cool LAB led to an effect on students' physics self-beliefs ($d_C = 0.4$), which is still at the hinge-point Hattie (2008) derived as indicator for educational relevance. Considering the short intervention time and that the improvement of students' self-beliefs was not the focus of the hands-on session, this result is promising and consistent with previous findings. Indeed, a meta-analysis by O'Mara et al. (2006) reports a mean effect size of $d_C = 0.5$ for the impact of self-concept interventions, potentially of much longer duration. Also, previous studies employing similar scales before and after similar hands-on sessions reported comparable or weaker results. For example, Molz (2016) reported almost the same trend with an average dispositional physics self-concept of 67% for his treatment groups and a situational physics self-concept of 75% (estimated average effect size over both treatment groups $d_C \approx 0.5$). In summary, S'Cool LAB performs very well compared to other hands-on out-of-school learning labs by triggering very high situational self-beliefs even for students whose dispositional physics self-beliefs were already very high.

Self-beliefs as particularly promising educational outcome

Compared to the related 'interest' construct, the regression models on 'self-beliefs' stood out for several reasons. First, the variance partition coefficient was much smaller for dispositional physics self-beliefs than for dispositional physics interest (15% vs. 28%), which indicates that group variables such as the teacher had a smaller effect on students' dispositional self-beliefs compared to their interest. Second, as the only significant novelty influence factor, dispositional physics self-beliefs only accounted for 18% of the variance in students' situational self-beliefs, which is much lower than the 31% of variance in their situational interest accounted for by novelty influence factors. This could indicate that students perceived the learning environment in S'Cool LAB differently from their physics classes. Indeed, Osborne et al. (2003) report "*the apparent contradiction between students' attitudes towards science in general and their attitudes towards school science*". In this case, students' perceptions of themselves in physics classes might be different, and to some degree independent, from their perception of themselves in physics activities outside school. Third, novelty experience factors alone accounted for half the variance in students' situational self-beliefs. This suggests that a favourable novelty experience at an out-of-school learning lab might lead directly to high situational self-beliefs, almost independent of students' dispositional self-beliefs. Considering the impact of self-beliefs

on course choices and career aspirations (Nagy et al., 2008; Parker et al., 2014), these results make the variable 'self-beliefs' particularly interesting as a desirable and feasible educational outcome of out-of-school learning opportunities.

S'Cool LAB closed the gender gap for interest and self-beliefs

Overall, female and male participants did not differ in their educational outcomes after taking part in half-day hands-on sessions in S'Cool LAB. However, girls benefited more from the hands-on sessions in S'Cool LAB with respect to their physics interest ($d_C = 0.8$) and self-beliefs ($d_{C,} = 0.5$) than did boys ($d_C = 0.5$ & 0.3). Consequently, the initial gender gap with girls reporting slightly lower dispositional physics interest ($d_C = 0.3$) and physics self-beliefs ($d_C = 0.4$) was closed with respect to their situational interest and self-beliefs. This result is consistent with previous findings, for example, by Burkam et al. (1997) and Engeln (2004). Moreover, in a large study with more than 11 000 students, Priemer et al. (2018) also report a negligible gender difference in students' situational interest after taking part in hands-on activities at out-of-school learning labs (estimated $d_C \approx 0.1$). While these effects on the 'gender gap' may not appear large, they refer to rather stable traits. Hence, it is remarkable that even a short half-day intervention can initiate a positive change.

Persistent and very high curiosity towards particles physics

When comparing students' epistemic curiosity towards particle physics before and after having taken part in hands-on S'Cool LAB workshops, no significant differences were identified. However, the level of students' curiosity towards particle physics was already very high in the pre-visit questionnaire (78% of the maximum possible score), which left little room for improvement towards the post-visit questionnaire (77% of the maximum possible score). When comparing these curiosity levels with levels for other topics, such as 48% for oscillations and waves (Hochberg, 2016), 62% for acoustics (Hirth, 2019), and 65% for radioactivity (Molz, 2016), particle physics stands out as an extremely promising topic to accommodate and maintain remarkably high levels of epistemic curiosity. This finding is also interesting in light of the results of the ROSE study (Sjøberg & Schreiner, 2010), which presented astrophysics as another modern physics topic that aroused high levels of epistemic interest. Here, other formal or informal learning opportunities could make use of students' desire to know more about modern physics by embedding other physics contents, for example, in a particle physics context. Indeed, particle physics can serve as a context for a variety of topics, ranging from the Lorentz force ensuring a circular particle trajectory in the LHC to horizontal projectile motion in the context of anti-hydrogen research at CERN.

Intertwined nature of educational outcomes confirmed

Many of the educational outcomes assessed in the framework of the S'Cool LAB studies are closely related theoretical constructs (see subsection 2.2.7). Consequently, most of the educational outcomes were significantly correlated with medium to high correlation coefficients (see section 7.5). In particular, all of the more affective outcomes (interest, enjoyment, helpfulness, and curiosity) were strongly correlated with each other and with students' perceived cognitive activation ($r > 0.5$). Overall, correlations between S'Cool LAB's educational outcomes showing their intertwined nature were consistent with previous studies, for example, by Cors (2016), and also consistent with the theoretical foundation of these variables (see section 2.2).

For example, the highest correlation was found between students' situational interest and curiosity towards particle physics ($r_{IE3,CS3} = 0.77, p = 2 \cdot 10^{-101}$), which supports the close connection between epistemic curiosity and interest (Alexander & Grossnickle, 2016). Moreover, students' situational interest and self-beliefs were strongly correlated with each other ($r_{IE3,SC3} = 0.54, p = 2 \cdot 10^{-40}$). This was expected since both can be interpreted as components of motivation. Here, Itzek-Greulich et al. (2017) reported a slightly lower correlation ($r = 0.45, p < 0.001$) in a similar out-of-school learning lab context. Students' conceptual understanding after taking part in a hands-on session in S'Cool LAB was only very weakly correlated with two other educational outcomes, namely with their curiosity towards particle physics ($r_{CS3,UN3} = 0.15, p = 0.002$) and their perception of easiness ($r_{EAS,UN3} = 0.18, p = 0.0001$). Here, previous studies reported slightly higher correlations between affective and cognitive outcomes. For example, Molz (2016) reported a correlation of $r = 0.3$ between achievement and motivation. However, based on a meta-analysis, Hansford and Hattie (1982) reported an average correlation between self-measures (such as self-concept or self-esteem) and achievement of $r = 0.21$, which confirms a similar value.

Cognitive activation as surprising variable in out-of-school learning

Very high levels of cognitive activation (78% of the scale's maximum) confirmed that students felt that S'Cool LAB did not only allow hands-on but also 'minds-on' manipulation. Moreover, cognitive activation seems to be at the intersection between the more affective outcomes and the more cognitive outcomes as it was strongly correlated with all affective outcomes but also with situational self-beliefs and, to a slightly lesser degree, with perceived easiness. Similarly, Simon (2019) reported perceived cognitive activation as strong predictor of students' interest in an out-of-school learning lab, and Engeln (2004) highlighted the importance of the comparable variable 'cognitive challenge'.

Still, the correlation between this rather cognitive variable and the more affective outcome variables is somewhat surprising, especially at an out-of-school learning opportunity. Here, educators usually focus on fun and impressive activities to catch students' interest. Moreover, only novelty experience factors accounted for significant amounts of variance of this educational outcome. Indeed, higher levels of perceived 'cognitive activation' were reached by students who perceived the out-of-school learning atmosphere as favourable with high levels of tinkering support ($\beta_{w/o} = 0.3^{***}$) and tutor support ($\beta_{w/o} = 0.3^{***}$) as well as high levels of setting orientation ($\beta_{w/o} = 0.2^{***}$). In summary, it seems to be particularly important to get students cognitively involved at out-of-school learning opportunities. Moreover, the level of cognitive activation students reach depends almost entirely of quality indicators of the learning opportunity. Consequently, cognitive activation is a desirable and feasible educational outcome of out-of-school learning labs.

Medium-sized effect on conceptual understanding

In the pre-visit questionnaire, students demonstrated a low conceptual understanding of key concepts relevant to the hands-on workshops in S'Cool LAB with an average score of 39%. After taking part in half-day hands-on session, it was still not easy for them to answer the selected concept test items, but their average score improved to 56%. Indeed, the most substantial effect of half-day hands-on sessions in S'Cool LAB was identified for students' conceptual understanding, which showed a medium-sized increase from the pre- to the post-visit questionnaire ($d_C = 0.7$, Hake's gain $g = 0.3$). Certainly, in many cases, the regular school setting may support learning processes better, and many out-of-school learning labs view learning rather as a 'side-effect', as they mainly focus on affective outcomes. However, half-day hands-on sessions in S'Cool LAB successfully demonstrate that out-of-school learning labs can significantly foster students' conceptual understanding without compromising the impact on other variables, such as their interest and self-beliefs. This results is consistent with finding by Molz (2016), who confirmed that even a short intervention with hands-on learning activities in an out-of-school learning lab about radioactivity can reduce the percentage of students answering according to well-known misconceptions about irradiation and contamination.

The design features of S'Cool LAB's activities proved successful

Previous studies suggested that out-of-school learning labs have a unique educational potential based on their special features, such as access to rare equipment (2.1.4), hands-on experimentation (2.1.6), and contact to scientists (2.1.7). However, the effectiveness of the activities depends on their design and their potential to engage students in cognitively activating activities while fostering their interest and self-beliefs. At the same time, it is

necessary to balance the amount of novelty participants experience to ensure enough cognitive capacities to successfully process the learning experiences. Therefore, the first steps of the S'Cool LAB project included the evidence-guided development of high-quality, hands-on activities. Consequently, the S'Cool LAB hands-on sessions were designed to provide an out-of-school learning lab with the best possible conditions for learning and development. The design features of S'Cool LAB's hands-on workshops were not studied directly within this PhD project. Instead, the workshop development followed well-established recommendations for optimising the educational potential of practical work.

For example, predict-observe-explain tasks were used to foster the links between hands-on and minds-on activities by activating students' theoretical reasoning skills. Also, the workshops in S'Cool LAB concentrated on a few selected but important tasks and key concepts. In contrast, tedious manipulations of high-tech equipment were omitted or well-guided to reduce extraneous cognitive load as much as possible and to focus students' cognitive capacity on germane cognitive load instead. In general, students reported low levels of cognitive load perceived during the workshops in S'Cool LAB. However, the questionnaire items related to the germane cognitive load of the different workshops were consistently rated highest among the different cognitive load components across all three workshops.

Together with very high values in students' perceived cognitive activation and with the good learning efficiency, this suggests that the workshop design as well as the guidance through process worksheets and by tutors were successful. Indeed, students could concentrate on the underlying physics principles instead of fighting with the equipment. Here, other out-of-school learning opportunities can build on these successful design elements, and also implement parts of the questionnaires to monitor key indicators of students' novelty experience.

9.1.2 Novelty influence factors and novelty experience factors

The role of novelty influence factors (student characteristics) and novelty experience factors (students' perception of laboratory characteristics) was studied systematically to gain insights into their predictive power for a wide range of educational outcomes and the 'mechanisms of action' of S'Cool LAB and out-of-school learning labs in general.

The S'Cool LAB studies presented suitable test instruments

The items that were included in the pre-visit and post-visit questionnaires of the S'Cool LAB main study showed suitable properties, the scales' dimensionalities were consistent with previous studies, and the selected subscales demonstrated good internal

consistency (7.3.14). Nevertheless, the distributions of the mean values of almost all scales deviated significantly from normal distributions as many scales were negatively skewed and included several extreme values in their heavy tails. This can be explained by the special sample of highly motivated and curious students that took part in the hands-on sessions in S'Cool LAB. In summary, the S'Cool LAB studies employed suitable test instruments to assess the variety of novelty factors and educational outcomes in a time-efficient way by using simple English questionnaire items with an international sample. Here, other out-of-school learning opportunities interested in investigating some of these factors or outcomes can benefit from the well-validated scales of the S'Cool LAB studies.

Educational outcomes were independent of age, gender, and prior experiences

Indeed, the control variables age and gender did not account for a significant amount of variance in any of the dependent variables. Also, students' prior experiences and English skills were not relevant for any of the linear regressions on the educational outcomes; the results suggest that the educational outcomes of S'Cool LAB were independent of students' English skills and their prior experiences. This result is a promising aspect for out-of-school learning opportunities, which also cater to a diverse audience from different countries and educational backgrounds. Indeed, (mostly European) students aged 16-19 were able to follow and benefit from hands-on activities in English.

Support variables were crucial for affective outcomes

The regression models on 'situational interest', 'curiosity towards particle physics', 'perceived enjoyment' and 'perceived helpfulness' were remarkably similar and confirmed a close relationship with the interest construct. Higher levels of these four educational outcomes were reached by students who were generally interested in physics and who were already curious about particle physics. Among the novelty experience factors, tinkering support (e.g., helpful instructions and easy to find materials) and tutor support (e.g., help in case of problems and display of fascination for physics) were the most relevant predictors. These variables consistently demonstrated high standardised regression coefficients ($\beta_{w/o} = 0.2 - 0.3$) for the more affective outcomes, namely situational interest, curiosity, perceived enjoyment and helpfulness as well as for the outcome 'cognitive activation'.

Hence, students' perception of support both by the learning environment and by the educators in the room were crucial elements of the novelty experience S'Cool LAB, which is consistent with previous findings. For example, Pawek (2009) identified the social factors 'support by tutors' and a positive 'learning atmosphere' as promising predictors for situational interest, and Simon (2019) highlighted the importance of clear instructions,

support by tutors, and a positive learning atmosphere. Consequently, other out-of-school learning opportunities should invest in recruiting and training their educators but also support learners by appropriately structuring their hands-on activities. Due to the popularity of inquiry-based learning, some educators falsely assume that open inquiry is the best possible activity type. However, the results of the S'Cool LAB studies suggest that guidance and scaffolding are highly valuable even for high-achieving students and when targeting affective outcomes.

Cognitive preparedness and cognitive load were crucial for cognitive outcomes

Setting orientation due to students' preparation (e.g., students knew what to expect in S'Cool LAB and were familiar with the schedule) was an important predictor of the more cognitive outcomes, namely situational self-beliefs, perceived easiness and cognitive activation with consistently high standardised regression coefficients ($\beta_{w/o} = 0.2$). Similarly, students' perceived cognitive load (e.g., students found the experiments difficult and the instructions hard to understand) was a significant predictor of the outcomes 'situational self-beliefs' ($\beta_{w/o} = -0.3$) and 'perceived easiness' ($\beta_{w/o} = -0.4$). Hence, students' perception of their (cognitive) preparedness and their cognitive load were vital components of the novelty experience in S'Cool LAB, in particular with respect to its cognitive effects. This is consistent with previous research results, for example, by Orion and Hofstein (1994), Streller (2015), and Molz (2016), who emphasised the importance of novelty-reducing preparation and the impact of cognitive load on learning and development. Consequently, out-of-school learning opportunities should invest in resources to help teachers prepare their students in advance to optimise their educational effectiveness.

Laboratory characteristics were more important than student characteristics

On average, novelty experience factors accounted, with about 40%, for twice the amount of variance in the educational outcomes than did novelty influence factors. This suggests that the characteristics of the out-of-school learning laboratory were more important than student characteristics with respect to students' educational outcomes. This is consistent with previous studies, for example, by Simon (2019), who reported that the support variables alone accounted for about 25% of the variance in students' interest.

Only for two outcomes, namely 'curiosity state' and 'conceptual understanding', did novelty influence factors account for a higher amount of variance than novelty experience factors. Here, students' curiosity towards particle physics was already very high in the pre-visit questionnaire and did not increase further after participation in the hands-on session in S'Cool LAB. Thus, it is plausible that the pre-visit curiosity state strongly predicts the post-

visit curiosity state. When considering students' conceptual understanding, it is well known that the strongest predictor of learning is what students already know (Ausubel, 1968; Hattie, 2008). Moreover, in both of these two cases, the items assessing the outcomes in the post-visit questionnaire were already included in a very similar way in the pre-visit questionnaire, which also explains the high amount of shared modelled variance between these two pairs of variables. In general, several novelty influence factors correlated with novelty experience factors, and their respective modelled variance showed an overlap. Therefore, novelty experience factors seem to mediate effects of several novelty influence factors as suggested by Cors (2016).

The novelty influence and experience factor framework proved very useful

Based on the results of the S'Cool LAB evaluation main study, it can be summarised that the novelty influence and experience factor framework described by Cors et al. (2017) is a very useful tool when modelling educational outcomes of an out-of-school learning opportunity: it not only provides a theoretical framework to consider student and laboratory characteristics in a systematic way but both types of novelty factors and their different dimensions accounted for high amounts of variance in several educational outcomes. On average, all novelty factors together accounted for about 50% of the variance in the educational outcomes, which indicated a good model fit. In particular, novelty factors affected educational outcomes most, if they belonged to similar dimensions. Indeed, novelty factors belonging to the cognitive dimension and factors of the dimension 'setting familiarity' were relevant mostly for cognitive outcomes. By contrast, novelty factors belonging to the affective dimension as well as the social and 'capabilities' dimension were mostly relevant for affective outcomes. Moreover, the novelty influence and experience factor framework helped to account for hierarchical data structures. Indeed, students taking part in the S'Cool LAB studies came in different groups with different teachers and shared similar novelty influence factors. Here, including student-level predictors accounted for a large proportion of the group-level variance. Consequently, when taking appropriate measures to account for inflated type I errors, a suitable selection of novelty influence (and experience) factors can support the use of single-level models on clustered data.

Modelling conceptual understanding presented a special case

Conceptual learning posed a special case when comparing the different regression models. Although several of the novelty influence and experience factors were weakly but significantly correlated with this outcome variable in a meaningful way (e.g., English skills, experience with particle physics, tinkering support and cognitive load), only predictors modelling what students already knew accounted for variance of their post-visit conceptual understanding. In particular, strong predictors of students' conceptual understanding after

taking part in the S'Cool LAB hands-on session were their initial conceptual understanding before coming to S'Cool LAB ($\beta_{w/o} = 0.4^{***}$) as well as their self-reported physics grade ($\beta_{w/o} = 0.2^{**}$), which is consistent with previous studies (Ausubel, 1968; Hattie, 2008).

Overall, these results suggest that conceptual learning gains were independent of students' perception of their novelty experience. This indicates that these learning gains are indeed more robust than more affective outcomes. Hence, other out-of-school learning opportunities should consider including conceptual understanding as one of their desired outcomes.

Multi-level models were better suited than single-level linear models

Students taking part in S'Cool LAB hands-on sessions had different teachers, came in different groups, and from different countries, which was reflected in variance partition coefficients of up to 31% with respect to the educational outcomes. Here, ignoring the hierarchical data structure translated into a dramatic overestimation of the effective sample size up to a factor of 11 and, consequently, severely downwardly biased p-values. Furthermore, even the predictors that were considered student-level variables, reflected the hierarchical data structure with variance partition coefficients between merely 5% (for curiosity trait) and 58% (for hands-on experience). In summary, students who came to S'Cool LAB on the same day as part of the same group shared similar novelty influence factors, for example their previous experience with hands-on experiments or their dispositional physics interest. Adding these novelty influence factors as student-level predictors into two-level models accounted for most of the group differences with respect to the educational outcomes. In contrast, novelty experience factors differed far less between different groups. This suggests that students from different groups perceived their novelty experience in S'Cool LAB in a similar way. Furthermore, group differences in students' educational outcomes can mostly be accounted for by group differences in students' novelty influence factors. Overall, in particular affective educational outcomes and students' conceptual understanding differed significantly among students coming in different groups, because the student groups differed significantly in their dispositional interest and their prior conceptual understanding.

In summary, the two-level models bore a striking resemblance to the respective single-level models and confirmed the same results. However, p-values were underestimated in single-level models on clustered data due to the overestimation of the effective sample size. Therefore, the two-level models were better suited for this data set than single-level models.

9.1.3 Students' conceptions about ionising radiation and radioactivity

Students participating in S'Cool LAB's hands-on particle physics workshops were also used to expand our current understanding of students' conceptions about ionising radiation and radioactivity as well as students' reasoning patterns while developing and validating a suitable quantitative test instrument. Here, a set of multi-true-false items was developed based on documented students' conceptions to measure and analyse students' conceptual understanding of radiation and radioactivity systematically and quantitatively.

Students' conceptions of radiation and radioactivity were reproduced

Only very few of the items generated based on previously documented students' conceptions were very easy for students taking part in S'Cool LAB workshops and, consequently, were answered correctly by at least 90% of the students. For example, very few students expected a patient or a copper object to 'glow in the dark' after being X-rayed for an hour. Indeed, this conception was identified by Boyes and Stanisstreet (1994), but, 26 years later, this conception seems to be no longer prevalent. Also, almost none of the students believed that the half-life is the time in which a radioactive source is dangerous, or that heat can turn stable atoms into radioactive atoms. Moreover, only a small fraction of the students answered that an imbalance of the number of electrons and neutrons causes atoms to be radioactive. Furthermore, very few students said that Wi-Fi radiation or visible light could ionise atoms. However, the majority of the items was sufficiently difficult for students taking part in S'Cool LAB workshops. Furthermore, eight items were very difficult, with success rates below the guessing rate of 50%. In particular, students struggled to identify the role of electrons in radioactive transformations and to distinguish phenomena of the nucleus from phenomena of the orbital. Instead of expecting an electron being emitted from the nucleus, more than half of the students answered that an electron drops to a lower energy level during one type of radioactive decay. Also, almost half of the students expected that an electron would be emitted from the orbital during radioactive decay. Furthermore, students confused irradiation and contamination, which might be based on a wrong model of radiation as something 'consisting of radioactive particles'. Consistently with this wrong model, an overwhelming number of students answered incorrectly that X-radiation accumulates inside the body if a patient is exposed to X-radiation for an hour.

These results suggest that students' conceptual understanding of radiation and radioactivity has not changed much over the past decades. Instead, most of the previously documented conceptions, which formed the basis of the 92 multiple-true-false items, were reproduced within this study.

Conceptions at the particle level and linguistic accuracy are crucial

Based on the source-radiation-absorber model, documented students' conceptions were categorised into six categories, which consisted of the three concepts 'source', 'radiation' and 'absorber' each on a sub-microscopic level and a macroscopic level. Consequently, the MTF items used in this study were developed to represent selected aspects for each of the six categories. Here, learners might struggle to link the two different levels similar to the level confusion described with respect to the 'chemistry triplet' (Johnstone, 1991). This level confusion might then even account for common misconceptions.

The results of the students' conception study confirmed that many misconceptions seem to be linked to the particle level. Consequently, educators need to pay special attention to these conceptions, when teaching radioactivity or topics that build upon the understanding of elementary particles and their interactions with matter. In particular, students seem to struggle with the development and use of an accurate model of radiation. Here, focusing on sub-microscopic processes, by introducing a particle model for all types of radiation and interactions of radiation with matter, might prove useful to avoid common misconceptions. Possibly, the term 'radiation' should be avoided altogether in favour of more specific descriptions, such as 'ionising particles'. These suggestions are consistent with the call for linguistic accuracy and the focus on the model aspect, especially when teaching elementary particle physics, by Wiener (2017). Moreover, educators should explicitly focus on the distinction between physics phenomena of the orbital and physics phenomena of the nucleus, for example, by highlighting their different energy scales. Finally, it is essential to focus learners' attention to the discrete and random nature of radioactive transformations, as suggested already by Jansky (2019).

MTF items revealed partial understanding independent of basic knowledge

In summary, students seemed to demonstrate neither an unstructured collection of phenomenological primitives as proposed by DiSessa (1988) nor a coherent understanding of scientific concepts. Instead, the results of the factor analysis suggest that, on the one hand, students argue rather consistently within certain areas, such as the macroscopic perspective on half-life. On the other hand, the results of the cluster analysis suggest that students' conceptual understanding of one area, such as half-life, is independent of their understanding of a different area, such as the sub-microscopic understanding of the radioactive decay. Most strikingly, these misconceptions seem to manifest themselves independently of the students' (prior) knowledge. Here, MTF items proved their potential to reveal this partial understanding by identifying even a mixture of accurate and inaccurate conceptions. Since severe misconceptions seem to occur independently of students' basic knowledge, it seems equally important to explicitly discuss

potential misconceptions and to teach the actual concepts and facts of a topic, such as radiation and radioactivity. Otherwise, pre-existing conceptions will remain unchallenged and might coexist with new conceptions.

Research on students' conceptions can be beneficial for out-of-school learning labs
In summary, the results of the study on students' conceptions confirm that it is worthwhile and possible to assess students' conceptual understanding of radiation and radioactivity quantitatively in the framework of out-of-school learning labs. Furthermore, MTF items provide a relatively easy way to do so, especially, if previous studies provided a list of relevant students' conceptions. This approach could also be beneficial for other content areas in which separate students' conceptions are known, but no concept test exists. Potentially, accompanying research about students' conceptions could help to maximise the cognitive outcomes of hands-on workshops. In particular, students may be surprised and cognitively activated when making observations that contradict their misconceptions. In this way, knowledge of students' conceptions can be used to make hands-on activities more challenging and cognitively activating. Finally, focusing on students' conceptions can place educators in out-of-school learning places in the role of science education researchers. This could not only help volunteering scientists to learn more from their teaching experience but also to highlight the importance of research-based interactions with learners of all ages.

9.2 Strengths and limitations

The strength of the presented S'Cool LAB evaluation study lies in the combination of an intervention in the form of particle physics hands-on sessions, which offer a very high educational potential, and a large international student sample. Furthermore, a systematically research-based assessment of a wide range of educational outcomes but also of numerous novelty influence and experience factors allowed to investigate the mechanism of action of S'Cool LAB in more detail and to take a broader perspective on out-of-school learning labs than previous studies. Moreover, the study of students' conceptions contributed to the few studies to measure and analyse students' conceptual understanding of radiation and radioactivity systematically and quantitatively.

Although S'Cool LAB at CERN offers a unique research setting and access to an interesting student sample, there are several limitations to the studies conducted within this PhD project. As it is usually the case for out-of-school learning opportunities, S'Cool LAB relies on the research sample provided by participants taking part in its offers. Therefore, students who took part in one of the S'Cool LAB studies were not randomly

selected. Instead, groups of students that were planning a school trip to CERN were selected based on the students' and their teacher's motivation. Here, many of the groups consisted of advanced level physic courses. Consequently, students who took part in the S'Cool LAB studies exhibited, for example, high amounts of interest in physics, high levels of general curiosity and high levels of physics self-beliefs. This bias is also reflected in the distributions of the scales' mean values with severe deviation from a normal distribution. Also, the S'Cool LAB study samples showed a gender bias with only about one-third female participants. Contrary to previous studies on hands-on out-of-school learning labs, students came from different, mostly European countries. Although the use of online questionnaires during the S'Cool LAB evaluation study proved successful and increased the time available for experimentation in S'Cool LAB, only about two-thirds of the students, who filled out the pre-visit questionnaire also filled out the post-visit questionnaire, which might have resulted in an additional bias in the sample towards students who had a positive experience in S'Cool LAB. Therefore, the results obtained by the S'Cool LAB studies should be applied with caution to the whole population of high-school students. However, the studies can be seen as a very successful proof-of-concept. Indeed, if hands-on sessions in S'Cool LAB have a positive impact on high-achieving students, they are likely to also be beneficial for average-achieving students. Indeed, the weak influence of prior knowledge on S'Cool LAB's educational outcomes suggests that the results of linear regression models may also be relevant for average students.

Similarly, if even high-achieving students display severe misconceptions about radiation and radioactivity, average-achieving students can be assumed to perform in a similar way or demonstrate misconceptions more frequently. Finally, most of the results of the S'Cool LAB studies are consistent with previous research results on out-of-school learning labs, which suggests that the results do not reflect the biased sample.

Moreover, all students filled out questionnaires in English, although for most of them, English was only their second or third language. While the language of the questionnaire items was kept as simple as possible, for example by relying on a vocabulary of a B1 level whenever possible, not all students might have understood the questionnaire items in the intended way. Furthermore, students' English reading comprehension skills might have affected how and how fast students answered questions about their conceptual understanding. Nevertheless, the good internal consistency of most scales suggests that language issues were not a major problem. Indeed, supported by the experience during hands-on sessions in S'Cool LAB, it can be concluded that most high-school students of the age range 16 to 19, who took part in the S'Cool LAB hands-on sessions, demonstrated sufficiently advanced English skills.

9.3 Conclusions and outlook for future research

The results presented in chapters 6, 7 and 8 generated several open questions, which could serve as a promising starting point for future research projects. In particular, the S'Cool LAB evaluation study confirmed positive results of the hands-on particle physics sessions in S'Cool LAB. Here, it would be interesting to compare the results with other topics or with different locations. Future research could examine the educational effectiveness of S'Cool LAB workshops on, for example, thermodynamics or study the effect of similar particle physics workshops at a more neutral location. Indeed, the 'CERN halo' could be responsible for some of the positive effects of S'Cool LAB. Also, it would be beneficial to replicate the S'Cool LAB evaluation study with a different student sample that could be recruited from schools in the local area to avoid the sample bias towards high-achieving and predominantly male students.

In general, high levels of situational interest are a promising starting point. However, the situational interest triggered by out-of-school learning labs typically fades within a few months and rarely stabilises into longer-lasting dispositional interest. For example, Mokhonko et al. (2014) report only very small effects on female learners' dispositional interest and dispositional self-beliefs in physics (Cohen's $d_C = 0.1 - 0.2$). To stabilise these effects, previous studies investigated the effectiveness of multiple visits (Guderian, 2007; Zehren, 2009) or embedding the out-of-school learning experience in school learning settings. Indeed, it would be worth studying the long-term effects of S'Cool LAB to examine whether the highly positive educational outcomes persist for a longer time. In doing so, one could even follow up on individual students who now pursue a STEM career or who even returned to CERN in a professional capacity.

S'Cool LAB was not able to foster students' curiosity because their curiosity towards particle physics was already very high before coming to CERN. Here, it would be interesting to study the origins of students' epistemic curiosity in more detail. Are students who are not planning a trip to CERN also curious about particle physics? Are students equally curious about other physics domains? And how does curiosity develop over time?

The S'Cool LAB evaluation study focused on the quantitative assessment of the mechanism of action in S'Cool LAB. However, additional qualitative research could complement these results. In particular, it would be interesting to study the importance of the strongest predictors, such as 'tutor support' or 'tinkering support' in more details. Furthermore, the results suggest that cognitive outcomes are affected more by cognitive novelty factors, whereas affective outcomes are affected more by affective and social

novelty factors. Here, future research could examine this relationship in more detail, for example, through manipulating specific novelty types and investigating the effect on different types of affective and cognitive outcomes.

The results of the linear models of the main study support the assumption mediating effects of novelty experience factors. Furthermore, the correlations between the different educational outcomes of S'Cool LAB confirmed their intertwined nature. Here, future studies could investigate causal relationships and model these, for example, through structural equation models.

The S'Cool LAB studies identified difficulties with the use of the word 'particle'. In particular, students showed problems to distinguish high-energy elementary particles, such as electrons, from particle systems, such as air molecules. Furthermore, the results of the study on students' conceptions suggest that several students consider radiation to consist of radioactive particles. Here, more research is needed to examine how particle models can be introduced consistently and accurately across different domains including chemistry, the structure of matter and radiation. Furthermore, future research projects should examine students' mental particle models across these different domains.

During this PhD project, MTF items were successfully developed and implemented to measure and analyse students' conceptions of radiation and radioactivity. Here, it would be beneficial to implement the use of MTF items in a different domain that provides similar challenges when it comes to assessment, for example quantum physics. Moreover, item response theory could complement the presented results. Also, a cluster analysis of the extracted factors allows categorising different types of students. Indeed, it would be interesting to perform cluster analyses on tests assessing students' conceptual understanding in different areas. The results of such analyses could then prove useful in presenting specific types of students with specific learning material. Moreover, a larger and more diverse sample could be used to validate the results obtained. However, more research is needed to determine to what extent the item clusters derived from factor analysis represent consistent students' reasoning within an area. Here, qualitative data would help to test the hypothesis that certain sub-microscopic misconceptions (e.g., a 'radioactive particles' conception of radiation) cause misconceptions at the macroscopic level (e.g., 'radiation accumulates inside the absorber').

Due to limited testing time, only some of the documented students' conceptions about radiation and radioactivity were adjusted to MTF items to represent selected concepts of the two-level sources-radiation-absorber models. Here, it would be desirable to include all concepts in a future test instrument to obtain an exhaustive overview of students'

conceptual understanding of radiation and radioactivity. Moreover, as a next step, the MTF items could be administered before and after an intervention to assess their sensitivity to learning.

Overall, the S'Cool LAB studies show that the topic of particle physics provides a unique opportunity to introduce high-school students to modern physics while taking their interest and curiosity into account without compromising conceptual understanding. However, more remains to be investigated and hopefully the results of this PhD project will lead to plenty and fruitful advancements in the field of particle physics education research.

10 Appendix

10.1 Pre-visit questionnaire of the pilot study (paper version, grey scale)

CERN S'Cool LAB

We would like to know you better

S'Cool LAB

Dear student,

You will soon visit CERN and work with hands-on particle physics experiments in S'Cool LAB. We are constantly working to improve our activities. For that, we need your help! With this first questionnaire, we would like to get to know you, your previous experiences and your interests.

Your answers will **remain anonymous** and **this is not a test**!

After you have worked with the experiments, you will receive a second questionnaire. So we can see how your answers before and after your S'Cool LAB day at CERN are connected, we need to create a unique identification code for you to use on each questionnaire. On the next page, we will ask some questions about you for this purpose.

Thank you very much for your help!

Contact:

CERN S'Cool LAB cern.ch/s-cool-lab
Julia Woithe | julia.woithe@cern.ch
CH-1211 Geneva 23

Identification code

Please mark the respective letter or digit like this:

A		C	...

1. What is the first letter of your first name?
(e.g. Julia → J)

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	...	

2. How many older sisters do you have?
(e.g. none → 0)

0	1	2	3	4	5	

3. In which month are you born?
(e.g. December → 12)

1	2	3	4	5	6
7	8	9	10	11	12

4. What is the first letter of the first name of your mother? *(e.g. Alexandra → A)*

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	...	

5. How many older brothers do you have?
(e.g. one → 1)

0	1	2	3	4	5	

How to fill out this questionnaire

Most of the time, you have to rate how much you agree or disagree with a given statement and select the box which corresponds best to your opinion.

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
Cats are very interesting animals.	☐	☐	☐	☐	☒	☐

If you want to change your answer or make a mistake, simply put a new cross in the right place and draw a circle around the new cross.

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
Cats are very interesting animals.	☐	☐	☐	☐	☒	⊗

If you want to comment on questions or answer in more detail, please use the text field at the very end of the questionnaire.

At this point we would already like to thank you for your help!
Enjoy filling out our questionnaire!

The S'Cool LAB team

Some information about you

De1	How old are you?	☐15 ☐16 ☐17	☐18 ☐19 ☐____
De2	You are …	☐ female	☐ male
De3	What grade are you in? / How many years have you been going to school? *(e.g. 11 years → grade 11)*		
De4	What is your native language?		
De5	What is your nationality?		
De6	In what country is your school?		
De7	What is the name of your school?		
De9	What is the name of your physics teacher?		
De10	When will you be visiting CERN?		
	For the following questions, please use the grading system that your teachers normally use.		
De11	What was your grade in Mathematics in your latest grade report?		
De12	What was your grade in Physics in your latest grade report?		
De13	What was your grade in English in your latest grade report?		

Among all the subjects you have ever studied in school - which are / were your most favourite / least favourite subjects?

Su1	Most favourite subject 1	
Su2	Most favourite subject 2	
Su3	Least favourite subject 1	
Su4	Least favourite subject 2	

How do you like your physics classes?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
IE1pr	I spend time on physics in my free time – even beyond homework.	☐	☐	☐	☐	☐	☐
IE2pr	I work harder in physics than in other school subjects.	☐	☐	☐	☐	☐	☐
IE3pr	I enjoy solving physics problems.	☐	☐	☐	☐	☐	☐
IE4pr	I discuss topics from my physics classes with my friends or my family.	☐	☐	☐	☐	☐	☐
IE5	Physics is one of my favourite school subjects.	☐	☐	☐	☐	☐	☐
IE6pr	At home, I use books or the internet to learn more about the topics from physics classes.	☐	☐	☐	☐	☐	☐
IE7	When I am working on physics problems, sometimes, I do not notice time passing.	☐	☐	☐	☐	☐	☐
IE8pr	I like our physics classes.	☐	☐	☐	☐	☐	☐
IE9pr	I would like to spend more time in school on physics.	☐	☐	☐	☐	☐	☐
IE10	I think the physics I learn at school is very different from “real” physics.	☐	☐	☐	☐	☐	☐

What do you think about your performance in physics and in doing experiments?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
sc1pr	Physics classes are fun.	☐	☐	☐	☐	☐	☐
sc2pr	I understand the learning material in physics.	☐	☐	☐	☐	☐	☐
sc3pr	I look forward to physics classes.	☐	☐	☐	☐	☐	☐
sc4	I expect that my performance in physics classes will be good in the future.	☐	☐	☐	☐	☐	☐
sc5pr	My performance in physics classes is good.	☐	☐	☐	☐	☐	☐
sc6pr	I actively take part in physics classes.	☐	☐	☐	☐	☐	☐
sc7pr	I find it easy to solve tasks in physics classes.	☐	☐	☐	☐	☐	☐
sc8pr	I am focused during physics classes.	☐	☐	☐	☐	☐	☐
sc9pr	Experiments help me to understand the learning material in physics.	☐	☐	☐	☐	☐	☐
sc10pr	I believe that my classmates think I am good in physics.	☐	☐	☐	☐	☐	☐
sce1pr	Physics experiments are fun.	☐	☐	☐	☐	☐	☐
sce2pr	Doing experiments in physics is easy for me.	☐	☐	☐	☐	☐	☐
sce3pr	I am good in doing physics experiments.	☐	☐	☐	☐	☐	☐
sce4pr	I find it easy to finish physics experiments in the time given.	☐	☐	☐	☐	☐	☐
sce5pr	I find it easy to solve tasks related to physics experiments.	☐	☐	☐	☐	☐	☐

How curious are you about new things?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
ct1	I find it fascinating to learn new things.	☐	☐	☐	☐	☐	☐
ct2	I am curious about new things or experiences.	☐	☐	☐	☐	☐	☐
ct3	Critical thinking about previous ideas is fun.	☐	☐	☐	☐	☐	☐
ct4	I like to learn things that are new to me.	☐	☐	☐	☐	☐	☐
ct5	I find new things inspiring.	☐	☐	☐	☐	☐	☐
ct6	When I hear about something new, I want to know more about it.	☐	☐	☐	☐	☐	☐
ct7	I enjoy studying things that I don't understand yet.	☐	☐	☐	☐	☐	☐
ct8	Thinking about new ideas and new concepts is fun.	☐	☐	☐	☐	☐	☐

Are you curious about particle physics?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
cs2	I want to learn more about particle physics.	☐	☐	☐	☐	☐	☐
cs3 pre	I feel a personal connection with particle physics.	☐	☐	☐	☐	☐	☐
cs8	I want to understand the topic of particle physics better.	☐	☐	☐	☐	☐	☐
cs9 pre	I find it fascinating to spend time on particle physics.	☐	☐	☐	☐	☐	☐

What are your experiences with (particle) physics?

Physics classes

pe1	How long (in years) have you been studying physics in school?	_____ years
pe2	How long (in minutes) per week do you spend in physics classes?	_____ minutes

Particle Physics

	How many hours did you spend on particle physics so far?	0 h	0-1 h	1-2 h	3-5 h	6-10 h	10-20 h	> 20 h
pe3	**during physics classes**	☐	☐	☐	☐	☐	☐	☐
pe4	**outside physics classes**	☐	☐	☐	☐	☐	☐	☐
pe5	**In total**	☐	☐	☐	☐	☐	☐	☐

What are your experiences with physics experiments and out-of-school learning places?

Experiments

	How often do the following activities occur in your physics class?	never	very rarely	once every six month	once every two month	once a month	every two weeks	every week
pe6	Demonstration (experiment presented by your teacher)	☐	☐	☐	☐	☐	☐	☐
pe7	Students working on experiments individually	☐	☐	☐	☐	☐	☐	☐
pe8	Groups of students working on experiments	☐	☐	☐	☐	☐	☐	☐
pe9	Videos, simulations or animations of experiments	☐	☐	☐	☐	☐	☐	☐

Out-of-school learning places

You can not only learn new things at school but also outside school - for example in museums and zoos.

	How often do you approx. visit the following places or use the following offers?	never	very rarely	Once every six month	Once every two month	Once a month	Every two weeks	Every week
peo1	Museum	☐	☐	☐	☐	☐	☐	☐
peo2	Zoo, aquarium or botanical garden	☐	☐	☐	☐	☐	☐	☐
peo3	Science centre or hands-on lab	☐	☐	☐	☐	☐	☐	☐
peo4	Scientific TV shows	☐	☐	☐	☐	☐	☐	☐
peo5	University or research laboratory	☐	☐	☐	☐	☐	☐	☐
peo6	Science websites	☐	☐	☐	☐	☐	☐	☐
peo7	Other: _____	☐	☐	☐	☐	☐	☐	☐

You did it! Thank you very much for your help!

If you have comments, please use the following page.

Do you have comments about specific questions or the questionnaire in general?

10.2 Post-visit questionnaire of the pilot study (paper version, grey scale)

CERN S'Cool LAB
We are interested in your opinion

S'Cool LAB

Dear student,

You recently visited CERN and worked with particle physics experiments in S'Cool LAB. We are constantly working to improve our activities. For that, we need your help! With this second questionnaire, we would like to know how you like S'Cool LAB and how you think about the way of working, the organisation and the learning atmosphere in S'Cool LAB.

Your answers will **remain anonymous** and **this is not a test**!

So we can see how your answers before and after your S'Cool LAB day at CERN are connected, we need to create a unique identification code for you to use on each questionnaire. On the next page, we will ask some questions about you for this purpose.

You will need about 30 minutes to complete the entire questionnaire. It will probably seem strange that some questions occur more than once, or are only slightly modified. However, this is normal for this type of questionnaire: it helps us make sure we understand your answers.

Thank you very much for your help!

Contact:

CERN S'Cool LAB cern.ch/s-cool-lab
Julia Woithe | julia.woithe@cern.ch
CH-1211 Geneva 23

Identification code

Please mark the respective letter or digit like this:

A	X	C	…

1. What is the first letter of your first name?
(e.g. Julia → J)

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	…	

2. How many older sisters do you have?
(e.g. none → 0)

0	1	2	3	4	5	

3. In which month are you born?
(e.g. December → 12)

1	2	3	4	5	6
7	8	9	10	11	12

4. What is the first letter of the first name of your mother? *(e.g. Alexandra → A)*

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	…	

5. How many older brothers do you have?
(e.g. one → 1)

0	1	2	3	4	5	

De10	When did you visit CERN's S'Cool LAB?	

How to fill out this questionnaire

Most of the time, you have to rate how much you agree or disagree with a given statement and select the box which corresponds best to your opinion.

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
Cats are very interesting animals.	☐	☐	☐	☐	☒	☐

If you want to change your answer or make a mistake, simply put a new cross in the right place and draw a circle around the new cross.

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
Cats are very interesting animals.	☐	☐	☐	☐	☒	⊗

There are some questions with a different scale:

Cats are …

very boring	☐	☐	☐	☐	☐	☐	☐	very interesting

If you think cats are very boring, please put a cross in the box at the very left.
If you think cats are very interesting, please put a cross in the box at the very right.
Use the other boxes to grade your opinion.

If you want to comment on questions or answer in more detail, please use the text field at the very end of the questionnaire.

At this point we would already like to thank you for your help!
Enjoy filling out our questionnaire!

The S'Cool LAB team

Evaluation – S'Cool LAB Day overall

How did you like your S'Cool LAB day at CERN?

s1 The **guided tour** at CERN

not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ very much

s2 Working on the **experiments in S'Cool LAB**

not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ very much

s3 The **whole day** at CERN including S'Cool LAB

not at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ very much

How do you feel about the durations?

s4 The **guided tour** at CERN

too short ☐ ☐ ☐ ☐ ☐ ☐ ☐ too long

s5 Working on the **experiments in S'Cool LAB**

too short ☐ ☐ ☐ ☐ ☐ ☐ ☐ too long

s6 The **whole day** at CERN including S'Cool LAB

too short ☐ ☐ ☐ ☐ ☐ ☐ ☐ too long

How much did you learn?

s7 During the **guided tour** at CERN

nothing at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ very much

s8 While working on the **experiments in S'Cool LAB**

nothing at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ very much

s9 During the **whole day** at CERN including S'Cool LAB

nothing at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ very much

What was the most interesting thing you have learned at CERN?

s10

Evaluation – Experiments in S'Cool LAB

In the following questions we ask you about your time in S'Cool LAB!

How did you like the experiments in S'Cool LAB?

sc **Cloud chamber**

not at all	☐	☐	☐	☐	☐	☐	☐	very much

se Experiments with the **electron tube**

not at all	☐	☐	☐	☐	☐	☐	☐	very much

sx Experiments with the **X-ray machine**

not at all	☐	☐	☐	☐	☐	☐	☐	very much

Mental stress during the time in S'Cool LAB

cl1 **How high was the level of mental stress** you experienced in S'Cool LAB?

very low	☐	☐	☐	☐	☐	☐	☐	very high

cl2 **How hard did you have to work** to perform the experiments?

very little	☐	☐	☐	☐	☐	☐	☐	very much

cl3 How **frustrated** were you during your time in S'Cool LAB?

very little	☐	☐	☐	☐	☐	☐	☐	very much

cl4 How do you rate the **duration** of the different experiments?

too short	☐	☐	☐	☐	☐	☐	☐	too long

cl5 How do you rate the **number** of different experiments you were working on?

not enough	☐	☐	☐	☐	☐	☐	☐	too many

Information about your working group in S'Cool LAB

Including yourself, who was working at your table?

		0	1	2	3	4	5	6
T1	How many students were in your working group?	☐	☐	☐	☐	☐	☐	☐
T2	How many members of your working group were female?	☐	☐	☐	☐	☐	☐	☐
T3	How many members of your working group were male?	☐	☐	☐	☐	☐	☐	☐

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
How much did you enjoy the time in S'Cool LAB?							
Sa1	I enjoyed using the equipment.	☐	☐	☐	☐	☐	☐
Sa2	I enjoyed preparing the experiments.	☐	☐	☐	☐	☐	☐
Sa3	I enjoyed doing the experiments.	☐	☐	☐	☐	☐	☐
Sa4	I enjoyed discussing the phenomena.	☐	☐	☐	☐	☐	☐
Sa5	I enjoyed the time in S'Cool LAB.	☐	☐	☐	☐	☐	☐
How easy was the work in S'Cool LAB for you?							
Sa6	It was easy to understand the instructions.	☐	☐	☐	☐	☐	☐
Sa7	It was easy to prepare the experiments.	☐	☐	☐	☐	☐	☐
sa8	It was easy to use the equipment.	☐	☐	☐	☐	☐	☐
Sa9	It was easy to get a result.	☐	☐	☐	☐	☐	☐
Sa10	It was easy to understand what it was all about.	☐	☐	☐	☐	☐	☐
Was your time in S'Cool LAB helpful for you?							
Sa11	The time in S'Cool LAB was helpful for my physics classes.	☐	☐	☐	☐	☐	☐
Sa12	The time in S'Cool LAB helped me to get better at doing experiments.	☐	☐	☐	☐	☐	☐
Sa13	The time in S'Cool LAB helped me to understand physics concepts.	☐	☐	☐	☐	☐	☐
Sa14	The time in S'Cool LAB helped me to get an idea about how scientists work.	☐	☐	☐	☐	☐	☐
Sa15	The time in S'Cool LAB helped me to get an idea about what CERN does.	☐	☐	☐	☐	☐	☐

Organisation and way of working in S'Cool LAB

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
top1	I was familiar with the schedule in S'Cool LAB.	☐	☐	☐	☐	☐	☐
top2	The time in S'Cool LAB was well organized.	☐	☐	☐	☐	☐	☐
top3	I was well prepared for my time in S'Cool LAB.	☐	☐	☐	☐	☐	☐
top4	I knew what to expect in S'Cool LAB.	☐	☐	☐	☐	☐	☐
top5	I was easily getting familiar with the different work places in S'Cool LAB.	☐	☐	☐	☐	☐	☐
top6	Before arriving at CERN, I knew how I was going to work in S'Cool LAB.	☐	☐	☐	☐	☐	☐
top7	The materials needed were easy to find.	☐	☐	☐	☐	☐	☐
wow 1	In my working group, we spent a lot of time on doing experiments on our own.	☐	☐	☐	☐	☐	☐
wow 2	We had enough time to work on the experiments on our own.	☐	☐	☐	☐	☐	☐
wow 3	In my working group, we spent a lot of time on listening to the explanations of the tutors.	☐	☐	☐	☐	☐	☐
wow 4	The tutors in S'Cool LAB spent too much time on explaining.	☐	☐	☐	☐	☐	☐
Tin1	It was hard for me to do the experiments.	☐	☐	☐	☐	☐	☐
Tin2	I was afraid to make mistakes when doing the experiments.	☐	☐	☐	☐	☐	☐
Tin3	It took a long time until I was familiar with the equipment and was able to start the experiments.	☐	☐	☐	☐	☐	☐
Tin4	It was fun to try out the experiments.	☐	☐	☐	☐	☐	☐
Tin5	I received enough information to perform the experiments.	☐	☐	☐	☐	☐	☐
Tin6	I was able to concentrate on the experiments, without "fighting" with the equipment.	☐	☐	☐	☐	☐	☐
Tin7	The instructions were helpful for doing the experiments.	☐	☐	☐	☐	☐	☐

Did you feel "cognitively activated" in S'Cool LAB?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
Ca1	I was focused when doing the experiments in S'Cool LAB.	☐	☐	☐	☐	☐	☐
Ca2	I was actively involved in doing the experiments in S'Cool LAB.	☐	☐	☐	☐	☐	☐
Ca3	I took a critical look at new ideas or material.	☐	☐	☐	☐	☐	☐
Ca4	I tried to distinguish between important and unimportant things when doing the experiments.	☐	☐	☐	☐	☐	☐
Ca5	I tried to connect what I was learning with things I already knew.	☐	☐	☐	☐	☐	☐
Ca6	I thought about what will happen during the experiments before starting.	☐	☐	☐	☐	☐	☐
Ca7	I was able to use my physics knowledge to interpret my observations during the experiments.	☐	☐	☐	☐	☐	☐

Did you feel overwhelmed in S'Cool LAB?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
Cle 1	I had problems interpreting my observations correctly.	☐	☐	☐	☐	☐	☐
Cle 2	The experiments were difficult.	☐	☐	☐	☐	☐	☐
Cle 3	I had problems understanding the physics principles of the experiments.	☐	☐	☐	☐	☐	☐
Cle 4	It was hard for me to understand the instructions.	☐	☐	☐	☐	☐	☐
Cle 5	I would have needed more assistance when working with the experiments.	☐	☐	☐	☐	☐	☐

How mentally demanding was the use of the equipment?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
	Cloud Chamber						
clc1	It was difficult to use the equipment.	☐	☐	☐	☐	☐	☐
clc2	It was difficult to use the equipment when doing the experiment.	☐	☐	☐	☐	☐	☐
clc3	I had problems to start working with the equipment.	☐	☐	☐	☐	☐	☐
clc4	I was overwhelmed by the features provided by the equipment.	☐	☐	☐	☐	☐	☐
clc5	I had problems with using the equipment to make observations.	☐	☐	☐	☐	☐	☐
	Experiments with the **electron tube**						
clel1	It was difficult to use the equipment.	☐	☐	☐	☐	☐	☐
clel 2	It was difficult to use the equipment when doing the experiment.	☐	☐	☐	☐	☐	☐
clel 3	I had problems to start working with the equipment.	☐	☐	☐	☐	☐	☐
clel 4	I was overwhelmed by the features provided by the equipment.	☐	☐	☐	☐	☐	☐
clel 5	I had problems with using the equipment to make observations.	☐	☐	☐	☐	☐	☐
	Experiments with the **X-ray machine**						
clx1	It was difficult to use the equipment.	☐	☐	☐	☐	☐	☐
clx2	It was difficult to use the equipment when doing the experiment.	☐	☐	☐	☐	☐	☐
clx3	I had problems to start working with the equipment.	☐	☐	☐	☐	☐	☐
clx4	I was overwhelmed by the features provided by the equipment.	☐	☐	☐	☐	☐	☐
clx5	I had problems with using the equipment to make observations.	☐	☐	☐	☐	☐	☐

What do you think about the tutors in S'Cool LAB?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
TS 1	I had the chance to ask the tutors in S'Cool LAB questions.	☐	☐	☐	☐	☐	☐
TS 2	I had the feeling that the tutors are fascinated by physics.	☐	☐	☐	☐	☐	☐
TS 3	The discussions with the tutors were important to me.	☐	☐	☐	☐	☐	☐
TS 4	I liked the working atmosphere in S'Cool LAB.	☐	☐	☐	☐	☐	☐
TS 5	The tutors responded to my questions and comments while doing experiments.	☐	☐	☐	☐	☐	☐
TS 6	The tutors helped me with problems while doing the experiments.	☐	☐	☐	☐	☐	☐
Ts mes	**Would you like to leave a message to the tutors in S'Cool LAB?**						

How did you like your time in S'Cool LAB?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
IE1po	I spent time on physics I learned about in S'Cool LAB in my free time – even beyond homework.	☐	☐	☐	☐	☐	☐
IE2po	During my time in S'Cool LAB I worked harder than normally in school.	☐	☐	☐	☐	☐	☐
IE3po	During my time in S'Cool LAB, I enjoyed solving physics problems.	☐	☐	☐	☐	☐	☐
IE4po	I discussed topics from S'Cool LAB with my friends or my family.	☐	☐	☐	☐	☐	☐
IE5	Physics is one of my favourite school subjects.	☐	☐	☐	☐	☐	☐
IE6po	At home, I used books or the Internet to learn more about the topics from S'Cool LAB.	☐	☐	☐	☐	☐	☐
IE7po	When I was working on physics problems in S'Cool LAB, I did not notice time passing.	☐	☐	☐	☐	☐	☐
IE8po	I liked my time in S'Cool LAB.	☐	☐	☐	☐	☐	☐
IE9po	I would like to spend more time in S'Cool LAB.	☐	☐	☐	☐	☐	☐
IE10po	Studying physics in S'Cool LAB was very different from my physics classes in school.	☐	☐	☐	☐	☐	☐

What do you think about your performance during your time in S'Cool LAB?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
sc1po	The time in S'Cool LAB was fun.	☐	☐	☐	☐	☐	☐
sc2po	I understood the learning material in S'Cool LAB.	☐	☐	☐	☐	☐	☐
sc3po	I looked forward to my time in S'Cool LAB.	☐	☐	☐	☐	☐	☐
sc4	I expect that my performance in physics will be good in the future.	☐	☐	☐	☐	☐	☐
sc5po	My performance in S'Cool LAB was good.	☐	☐	☐	☐	☐	☐
sc6po	I actively took part in doing experiments in S'Cool LAB.	☐	☐	☐	☐	☐	☐
sc7po	I found it easy to solve tasks in S'Cool LAB.	☐	☐	☐	☐	☐	☐
sc8po	I was focused during my time in S'Cool LAB.	☐	☐	☐	☐	☐	☐
sc9po	The experiments in S'Cool LAB helped me to understand the learning material.	☐	☐	☐	☐	☐	☐
sc10po	I believe that during my time in S'Cool LAB my classmates thought I am good in physics.	☐	☐	☐	☐	☐	☐
sce1po	The physics experiments in S'Cool LAB were fun.	☐	☐	☐	☐	☐	☐
sce2po	Doing experiments in S'Cool LAB was easy for me.	☐	☐	☐	☐	☐	☐
sce3po	I was good at doing experiments in S'Cool LAB.	☐	☐	☐	☐	☐	☐
sce4po	During my time in S'Cool LAB, I found it easy to finish in the time given.	☐	☐	☐	☐	☐	☐
sce5po	I found it easy to solve tasks related to physics experiments in S'Cool LAB.	☐	☐	☐	☐	☐	☐

Are you curious about particle physics?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
CSP 1	My time in S'Cool LAB drew my attention to the topic of particle physics.	☐	☐	☐	☐	☐	☐
CSP 2	I want to learn more about particle physics.	☐	☐	☐	☐	☐	☐
CSP 3	I felt a personal connection with the activities in S'Cool LAB.	☐	☐	☐	☐	☐	☐
CSP 4	My time in S'Cool LAB allowed me to study phenomena I didn't know before.	☐	☐	☐	☐	☐	☐
CSP 5	The time in S'Cool LAB made me curious to learn more about particle physics.	☐	☐	☐	☐	☐	☐
CSP 6	The particle physics experiments in S'Cool LAB were inspiring to me.	☐	☐	☐	☐	☐	☐
CSP 7	I am more curious about particle physics now, after participating in S'Cool LAB activities.	☐	☐	☐	☐	☐	☐
CSP 8	I want to understand the topic of particle physics better.	☐	☐	☐	☐	☐	☐
CSP 9	I find it fascinating to spend time on particle physics.	☐	☐	☐	☐	☐	☐
CSP 10	Through the use of technology, the experiments illustrated phenomena that would otherwise be impossible to explore on my own.	☐	☐	☐	☐	☐	☐

How did you like the laboratory?

		very bad	bad	rather bad	rather good	good	very good
R1	Room in general	☐	☐	☐	☐	☐	☐
R2	Room temperature	☐	☐	☐	☐	☐	☐
R3	Quality of the air	☐	☐	☐	☐	☐	☐
R4	Decoration	☐	☐	☐	☐	☐	☐
R5	Cleanliness	☐	☐	☐	☐	☐	☐
R6	Furniture	☐	☐	☐	☐	☐	☐

How important were the following aspects of your trip to CERN for you?

		not at all important	not important	rather not important	rather important	important	very important
I1	Doing experiments on my own	☐	☐	☐	☐	☐	☐
I2	Fun	☐	☐	☐	☐	☐	☐
I3	No school	☐	☐	☐	☐	☐	☐
I4	Learning more about physics	☐	☐	☐	☐	☐	☐
I5	Learning more about particle physics	☐	☐	☐	☐	☐	☐
I6	Having visited S'Cool LAB	☐	☐	☐	☐	☐	☐
I7	Having visited CERN	☐	☐	☐	☐	☐	☐

Some last questions …

Please rate the following statements.

		completely disagree	disagree	disagree a little	agree a little	agree	completely agree
e1	During the guided tour, I was able to make connections to my previous physics knowledge.	☐	☐	☐	☐	☐	☐
e2	While doing the experiments, I was able to make connections to my previous physics knowledge.	☐	☐	☐	☐	☐	☐
e3	While doing the experiments, I was able to make connections to the content of the guided tour.	☐	☐	☐	☐	☐	☐
e4	I will recommend S'Cool LAB to my friends.	☐	☐	☐	☐	☐	☐

e5	Did you have to deal with foreign languages in S'Cool LAB?	☐ yes	☐ no

If yes:

		completely disagree	disagree	disagree a little	agree a little	agree	completely agree
e6	The foreign languages used in S'Cool LAB made it difficult for me to follow the activities.	☐	☐	☐	☐	☐	☐
e7	It would have been much easier for me, if all activities in S'Cool LAB would take place in my native language.	☐	☐	☐	☐	☐	☐

Feedback about the questionnaire

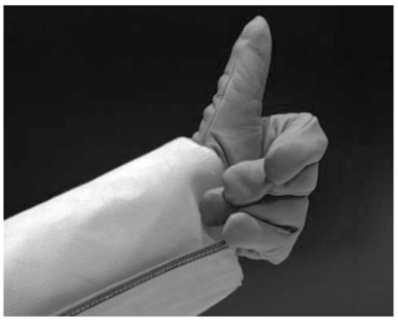
e8 What do you think about the length of this questionnaire?

very short ☐ ☐ ☐ ☐ ☐ ☐ ☐ very long

e9 How difficult was it for you to understand the questions in this questionnaire?

very difficult ☐ ☐ ☐ ☐ ☐ ☐ ☐ very easy

You did it! Thank you very much for your help!



Do you have comments about specific questions or the questionnaire in general?

_____

_____

_____

_____

_____

_____

_____

Is there anything else you wanted to tell us?

_____

_____

_____

_____

10.3 Pre-visit questionnaire of the main study (online version, grey scale)

We would like to get to know you!

Page 1

How to fill out this questionnaire

1 * Most of the time, you have to rate how much you agree or disagree with a given statement and select the field which corresponds best to your opinion. Below you will find an example. (Why cats? Because we like cats!)

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
Cats are very interesting animals.	○	○	○	○	○	○

If you want to comment on questions or answer in more detail, please use the text field at the very end of the questionnaire. At this point we would already like to thank you for your help! Enjoy filling out our questionnaire!

The S'Cool LAB team

Page 2

Identification code

After you have worked with the experiments, you will receive a second questionnaire. So we can see how your answers before and after your S'Cool LAB day at CERN are connected, we need to create a unique identification code for you to use on each questionnaire. Therefore, we will ask some questions about you for this purpose.

2 * What is the first letter of your first name? (e.g. Julia → J)

Choose...

3 * How many older sisters do you have? (e.g. none → 0)

Choose...

4 * In which month are you born? (e.g. December → 12)

Choose...

5 * What is the first letter of the first name of your mother? (e.g. Alexandra → A)

Choose...

6 * How many older brothers do you have? (e.g. one → 1)

Choose...

Page 3

Some information about you

7 * How old are you?

8 * You are ...

○ female ○ male

9 * What grade are you in? / How many years have you been going to school? (e.g. 11 years → grade 11)

10 * What is your native language?

Choose...

11 What is your native language?

Page 5

12 * What is your nationality?

Choose...

Page 6

13 What is your nationality?

Page 7

14 * In what country is your school?

Choose...

Page 8

15 * In what country is your school?

Page 9

16 * What is the name of your physics teacher?

Page 10

17 * For the following questions, please use the grading system that your teachers normally use.
What was your grade in Mathematics in your latest grade report?

18 * What was your grade in Physics in your latest grade report?

19 * What was your grade in English in your latest grade report?

Page 11

Among all the subjects you have ever studied in school - which are / were your most favourite subjects?

Please select 2!

20 * most favourite subject 1

○ Art ○ Astronomy ○ Biology ○ Chemistry ○ Computer science ○ Economics
○ English ○ Ethics ○ Foreign language ○ Geography / Geology ○ History
○ Literature ○ Mathematics ○ Music ○ Physical Education (P.E.) / Sports ○ Physics
○ Politics ○ Psychology ○ Religion ○ Science ○ Technology
○ Other:

21 * most favourite subject 2

○ Art ○ Astronomy ○ Biology ○ Chemistry ○ Computer science ○ Economics
○ English ○ Ethics ○ Foreign language ○ Geography / Geology ○ History
○ Literature ○ Mathematics ○ Music ○ Physical Education (P.E.) / Sports ○ Physics
○ Politics ○ Psychology ○ Religion ○ Science ○ Technology
○ Other:

Among all subjects you have ever studied in school - which are / were your least favourite subjects?

Please select 2!

22 * least favourite subject 1

○ Art ○ Astronomy ○ Biology ○ Chemistry ○ Computer science ○ Economics
○ English ○ Ethics ○ Foreign language ○ Geography / Geology ○ History
○ Literature ○ Mathematics ○ Music ○ Physical Education (P.E.) / Sports ○ Physics
○ Politics ○ Psychology ○ Religion ○ Science ○ Technology
○ Other:

23 * least favourite subject 2

○ Art ○ Astronomy ○ Biology ○ Chemistry ○ Computer science ○ Economics
○ English ○ Ethics ○ Foreign language ○ Geography / Geology ○ History
○ Literature ○ Mathematics ○ Music ○ Physical Education (P.E.) / Sports ○ Physics
○ Politics ○ Psychology ○ Religion ○ Science ○ Technology
○ Other:

24 * How do you like your physics classes?

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
I spend time on physics in my free time – even beyond homework.	○	○	○	○	○	○
I enjoy solving physics problems.	○	○	○	○	○	○
I discuss topics from my physics classes with my friends or my family.	○	○	○	○	○	○
When I am working on physics problems, sometimes, I do not notice time passing.	○	○	○	○	○	○
I like our physics classes.	○	○	○	○	○	○
I would like to spend more time in school on physics.	○	○	○	○	○	○

Page 13

25 * What do you think about your performance in physics and in doing experiments?

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree	N/A
I understand the learning material in physics.	○	○	○	○	○	○	○
I look forward to physics classes.	○	○	○	○	○	○	○
My performance in physics classes is good.	○	○	○	○	○	○	○
I find it easy to solve tasks in physics classes.	○	○	○	○	○	○	○
I believe that my classmates think I am good in physics.	○	○	○	○	○	○	○
Doing experiments in physics is easy for me.	○	○	○	○	○	○	○
I am good in doing physics experiments.	○	○	○	○	○	○	○
I find it easy to finish physics experiments in the time given.	○	○	○	○	○	○	○
Experiments help me to understand the learning material in physics.	○	○	○	○	○	○	○

Page 14

26 * How curious are you about new things?

Please rate the following statements

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
I find it fascinating to learn new things.	○	○	○	○	○	○
I am curious about new things or experiences.	○	○	○	○	○	○
I like to learn things that are new to me.	○	○	○	○	○	○
I find new things inspiring.	○	○	○	○	○	○
When I hear about something new, I want to know more about it.	○	○	○	○	○	○

27 * Are you curious about particle physics?

Please rate the following statements

	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
I want to learn more about particle physics.	○	○	○	○	○	○
I feel a personal connection with particle physics.	○	○	○	○	○	○
I want to understand the topic of particle physics better.	○	○	○	○	○	○
I find it fascinating to spend time on particle physics.	○	○	○	○	○	○

Page 16

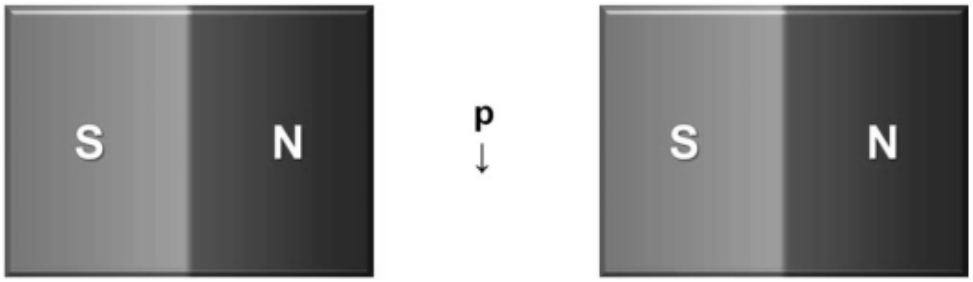
Some questions about physics

You will not know the answer to all of the following questions yet. However, it is extremely important that you fill in the questionnaire **without help from your teacher or your classmates** because we are interested in your ideas!

If you are not sure about one of the questions, just guess and select the answer option which corresponds best to your opinion. **This is not a test**! You will not be judged for your answers in any way.

Page 17

28 * A proton is moving downwards at constant speed through the uniform magnetic field between two bar magnets.



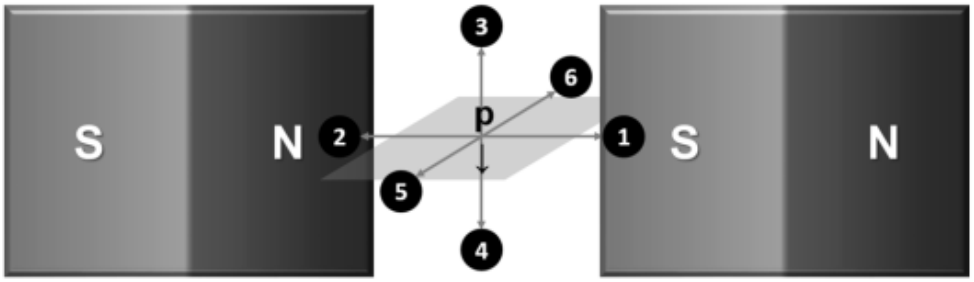
Do the magnets exert a force on the proton?

- ○ Yes, the magnets exert a force on the proton.
- ○ No, the magnets do not exert a force on the proton.

Page 18

29 * If the magnets exert a force on the proton: In which direction does the force point?

The coordinate system below might help you to identify the direction.



- ○ to the right (1)
- ○ to the left (2)
- ○ upwards (3)
- ○ downwards (4)
- ○ out of the page (5)
- ○ into the page (6)
- ○ in a different direction

Page 19

30 * Explanation:

- ○ The south pole S of the magnets attracts the proton.
- ○ The north pole N of the magnets attracts the proton.
- ○ The magnetic force acts in the direction of the magnetic field.
- ○ The magnetic force acts in the direction of the movement of the proton.
- ○ The magnetic force acts on the proton according to the right-hand-rule.
- ○ None of the explanations above apply.

31 * How confident are you that your answers are correct?

not confident at all ○ ○ ○ ○ ○ very confident

Page 20

32 * A proton is moving to the right at constant speed through the uniform magnetic field between two bar magnets.

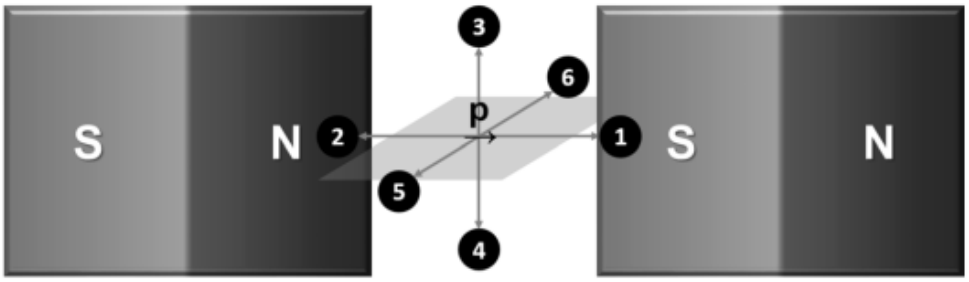
Do the magnets exert a force on the proton?

- ○ Yes, the magnets exert a force on the proton.
- ○ No, the magnets do not exert a force on the proton.

Page 21

33 * If the magnets exert a force on the proton: In which direction does the force point?

The coordinate system below might help you to identify the direction.



- ○ to the right (1)
- ○ to the left (2)
- ○ upwards (3)
- ○ downwards (4)
- ○ out of the page (5)
- ○ into the page (6)
- ○ in a different direction

Page 22

34 * Explanation:

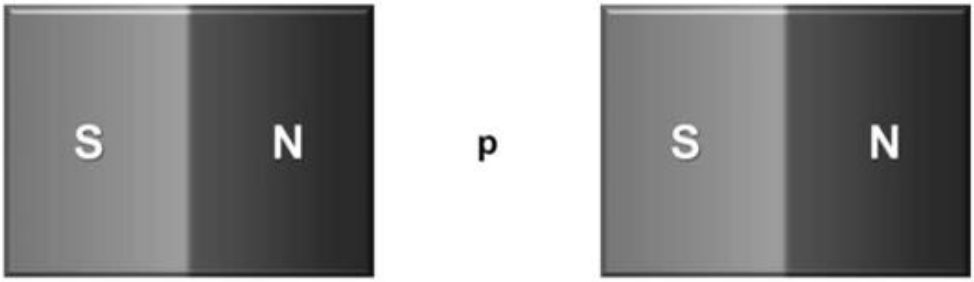
- ○ The south pole S of the magnets attracts the proton.
- ○ The north pole N of the magnets attracts the proton.
- ○ The magnetic force acts in the direction of the magnetic field.
- ○ The magnetic force acts in the direction of the movement of the proton.
- ○ The magnetic force acts on the proton according to the right-hand-rule.
- ○ None of the explanations above apply.

35 * How confident are you that your answers are correct?

not confident at all ○ ○ ○ ○ ○ very confident

Page 23

36 * A proton is at rest in the uniform magnetic field between two bar magnets.



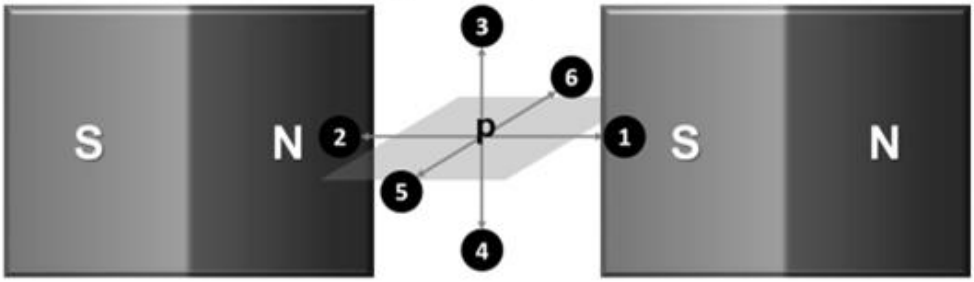
Do the magnets exert a force on the proton?

- ○ Yes, the magnets exert a force on the proton.
- ○ No, the magnets do not exert a force on the proton.

Page 24

37 * If the magnets exert a force on the proton: In which direction does the force point?

The coordinate system below might help you to identify the direction.



- ○ to the right (1)
- ○ to the left (2)
- ○ upwards (3)
- ○ downwards (4)
- ○ out of the page (5)
- ○ into the page (6)
- ○ in a different direction

Page 25

38 * Explanation:

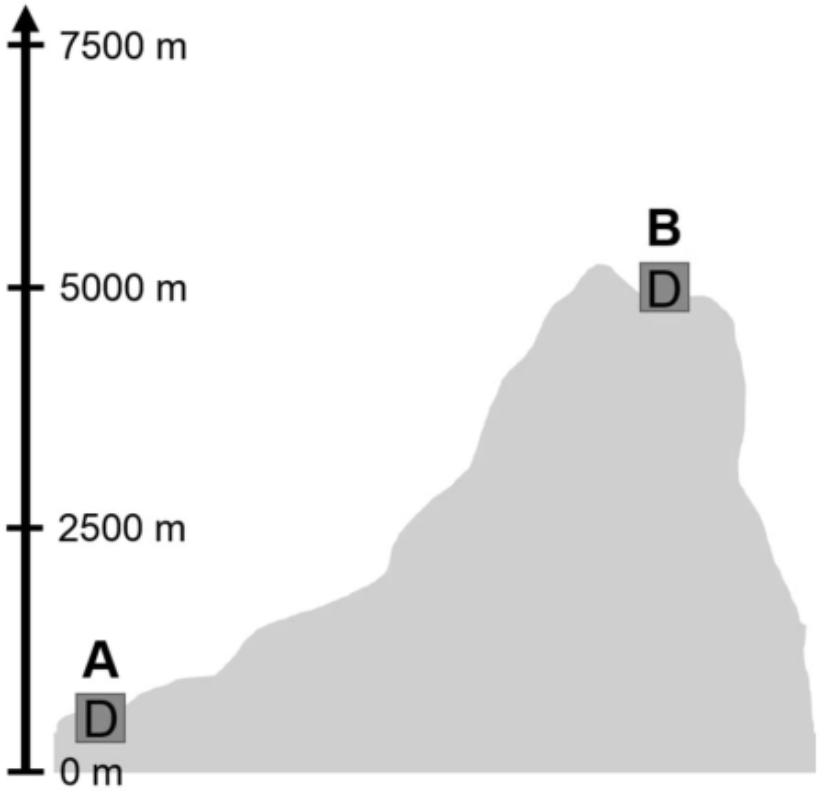
- ○ The south pole S of the magnets attracts the proton.
- ○ The north pole N of the magnets attracts the proton.
- ○ The magnetic force acts in the direction of the magnetic field.
- ○ The magnetic force acts in the direction of the movement of the proton.
- ○ The magnetic force acts on the proton according to the right-hand-rule.
- ○ None of the explanations above apply.

39 * How confident are you with your answers?

not confident at all ○ ○ ○ ○ ○ very confident

Page 26

40 * Imagine you compare two particle detectors [D]. One is placed in the valley at 0 m altitude. The other one is placed on top of a mountain at 5000 m altitude.



Will the particle detector on top of the mountain detect a higher number of particles than the detector at 0 m?

- ○ Yes, the particle detector on top of the mountain will detect a HIGHER number of particles.
- ○ No, the particle detector on top of the mountain will detect a LOWER number of particles.
- ○ No, there won't be a difference.

41 * Explanation:

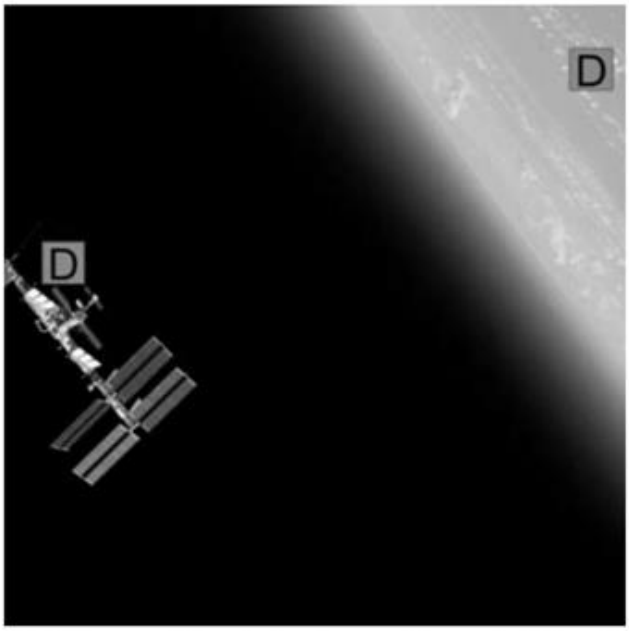
- ○ The properties of the atmosphere on top of the mountain and in the valley are the same.
- ○ The properties of the ground on top of the mountain and in the valley are the same.
- ○ The Earth is the main source of the detected particles.
- ○ Space is the main source of the detected particles.
- ○ The atmosphere has a shielding effect.
- ○ There are fewer air molecules on top of the mountain.
- ○ different explanation:

42 * How confident are you that your answers are correct?

not confident at all ○ ○ ○ ○ ○ very confident

Page 27

43 * Imagine you compare two particle detectors [D]. One is placed on the surface of the Earth at 0 km altitude. The other one is placed on a satellite at 400 km altitude.



Will the particle detector on the satellite detect a higher number of particles than the detector on the surface of the Earth?

- ○ Yes, the particle detector on the satellite will detect a HIGHER number of particles.
- ○ No, the particle detector on the satellite will detect a LOWER number of particles.
- ○ No, there won't be a difference.

44 * Explanation:

- ○ The properties of the atmosphere at 400 km altitude and at 0 km altitude are the same.
- ○ The Earth is the main source of the detected particles.
- ○ Space is the main source of the detected particles.
- ○ The atmosphere has a shielding effect.
- ○ There are fewer air molecules in space.
- ○ different explanation:

45 * How confident are you that your answers are correct?

not confident at all ○ ○ ○ ○ ○ very confident

Page 28

46 * Is it dangerous for the cleaning staff of a hospital when they clean a room with an X-ray machine?

- ○ Yes, parts of the X-ray device are exposed to X-rays and therefore become radioactive themselves.
- ○ Yes, because X-ray machines are always a source of radiation, even if they are switched off.
- ○ No, there is no danger, because the X-ray units are switched off during the cleaning.
- ○ No, provided that the entire air of the treatment rooms has been exchanged before the cleaning.
- ○ No, provided that the cleaning staff waits a while before they enter the room.

47 * How confident are you that your answer is correct?

not confident at all ○ ○ ○ ○ ○ very confident

Page 29

48 * A watch is X-rayed for 1 hour inside an X-ray machine. Is it dangerous to wear the watch afterwards?

○ Yes, ionizing radiation accumulates in the watch, therefore the watch will emit radiation afterwards.

○ Yes, the watch has been exposed to X-rays and therefore the atoms in the watch become radioactive.

○ No, to become radioactive itself, the watch needs to be exposed to X-rays for a longer time.

○ No, as long as no radioactive atoms are transferred to the watch, the watch does not emit radiation.

49 * How confident are you that your answer is correct?

not confident at all ○ ○ ○ ○ ○ very confident

Page 30

What are your experiences with (particle) physics?

Physics classes

50 * How long (in years) have you been studying physics in school?

51 * How long (in minutes) per week do you spend in physics classes?

52 * **Particle Physics**

How many hours did you spend on particle physics so far?

	0 h	0-1 h	1-2 h	3-5 h	6-10 h	10-20 h	> 20 h
during physics classes	○	○	○	○	○	○	○
outside physics classes	○	○	○	○	○	○	○
in total	○	○	○	○	○	○	○

Page 31

What are your experiences with physics experiments and out-of-school learning places?

Page 32

53 * **Experiments**

How often do the following activities occur in your physics class?

	never	very rarely	once every six months	once every two months	once a month	every two weeks	every week
Demonstration (experiment presented by your teacher)	○	○	○	○	○	○	○
Students working on experiments individually	○	○	○	○	○	○	○
Groups of students working on experiments	○	○	○	○	○	○	○
Videos, simulations or animations of experiments	○	○	○	○	○	○	○

54 * **Out-of-school learning places**

You can not only learn new things at school but also outside school - for example in museums and zoos. How often do you approx. visit the following places or use the following offers?

	never	very rarely	once every six month	once every two month	once a month	every two weeks	every week
Museum	○	○	○	○	○	○	○
Zoo, aquarium or botanical garden	○	○	○	○	○	○	○
Science centre or hands-on lab	○	○	○	○	○	○	○
Scientific TV shows	○	○	○	○	○	○	○
University or research laboratory	○	○	○	○	○	○	○
Science websites	○	○	○	○	○	○	○
Other offers	○	○	○	○	○	○	○

55 Other offers of out-of-school learning places you use regularily:

56 Do you have comments about specific questions or the questionnaire in general?

10.4 Inter-visit questionnaire of the main study (paper version, grey scale)

Identification code

Please mark the respective letter or digit like this:

A	X	C	…

1. What is the first letter of your first name?
(e.g. Julia → J)

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	…	

2. How many older sisters do you have?
(e.g. none → 0)

0	1	2	3	4	5	

3. In which month are you born?
(e.g. December → 12)

1	2	3	4	5	6
7	8	9	10	11	12

4. What is the first letter of the first name of your mother? *(e.g. Alexandra → A)*

A	B	C	D	E	F	G
H	I	J	K	L	M	N
O	P	Q	R	S	T	U
V	W	X	Y	Z	…	

5. How many older brothers do you have?
(e.g. one → 1)

0	1	2	3	4	5	

Are you curious about particle physics?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
cs2	I want to learn more about particle physics.	☐	☐	☐	☐	☐	☐
cs8	I want to understand the topic of particle physics better.	☐	☐	☐	☐	☐	☐
cs9 pre	I find it fascinating to spend time on particle physics.	☐	☐	☐	☐	☐	☐

How do you like your physics classes?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
IE1pr	I spend time on physics in my free time – even beyond homework.	☐	☐	☐	☐	☐	☐
IE3pr	I enjoy solving physics problems.	☐	☐	☐	☐	☐	☐
IE4pr	I discuss topics from my physics classes with my friends or my family.	☐	☐	☐	☐	☐	☐
IE7	When I am working on physics problems, sometimes, I do not notice time passing.	☐	☐	☐	☐	☐	☐
IE8pr	I like our physics classes.	☐	☐	☐	☐	☐	☐
IE9pr	I would like to spend more time in school on physics.	☐	☐	☐	☐	☐	☐

What do you think about your performance in physics and in doing experiments?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
sc2pr	I understand the learning material in physics.	☐	☐	☐	☐	☐	☐
sc3pr	I look forward to physics classes.	☐	☐	☐	☐	☐	☐
sc5pr	My performance in physics classes is good.	☐	☐	☐	☐	☐	☐
sc7pr	I find it easy to solve tasks in physics classes.	☐	☐	☐	☐	☐	☐
sc10pr	I believe that my classmates think I am good in physics.	☐	☐	☐	☐	☐	☐
sc9pr	Experiments help me to understand the learning material in physics.	☐	☐	☐	☐	☐	☐
sce2pr	Doing experiments in physics is easy for me.	☐	☐	☐	☐	☐	☐
sce3pr	I am good in doing physics experiments.	☐	☐	☐	☐	☐	☐
sce4pr	I find it easy to finish physics experiments in the time given.	☐	☐	☐	☐	☐	☐

10.5 Post-visit questionnaire of the main study (online version, grey scale)

We are interested in your opinion!

Page 1

1 * What is the name of your school?

Choose...

Page 2

Identification code

So we can see how your answers before and after your S'Cool LAB day at CERN are connected, we need to create a unique identification code for you to use on each questionnaire. Therefore, we will ask again some questions about you for this purpose.

2 * **What is the first letter of your first name? *(e.g. Julia → J)***

Choose...

3 * **How many older sisters do you have? *(e.g. none → 0)***

Choose...

4 * **In which month are you born? *(e.g. December → 12)***

Choose...

5 * **What is the first letter of the first name of your mother? *(e.g. Alexandra → A)***

Choose...

6 * **How many older brothers do you have? *(e.g. one → 1)***

Choose...

Page 3

Evaluation – S'Cool LAB Day overall

How did you like your S'Cool LAB day at CERN?

7 * The **guided tour** at CERN

not at all ○ ○ ○ ○ ○ ○ ○ very much

8 * Working on the **experiments in S'Cool LAB**

not at all ○ ○ ○ ○ ○ ○ ○ very much

9 * The **whole day** at CERN including S'Cool LAB

not at all ○ ○ ○ ○ ○ ○ ○ very much

How do you feel about the durations?

10 * The **guided tour** at CERN

too short ○ ○ ○ ○ ○ ○ ○ too long

11 * Working on the **experiments in S'Cool LAB**

too short ○ ○ ○ ○ ○ ○ ○ too long

12 * The **whole day** at CERN including S'Cool LAB

too short ○ ○ ○ ○ ○ ○ ○ too long

13 * We had enough time to work on the experiments on our own **in S'Cool LAB**.

completely disagree ○ ○ ○ ○ ○ ○ ○ completely agree

How much did you learn?

14 * During the **guided tour** at CERN

nothing at all ○ ○ ○ ○ ○ ○ ○ very much

15 * While working on the **experiments in S'Cool LAB**

nothing at all ○ ○ ○ ○ ○ ○ ○ very much

16 * During the **whole day** at CERN including S'Cool LAB

nothing at all ○ ○ ○ ○ ○ ○ ○ very much

17 * **What was the most interesting thing you have learned at CERN?**

Paragraph

Path: p

Page 4

Evaluation – Experiments in S'Cool LAB

In the following questions we ask you about your time in S'Cool LAB!

On which experiments did you work?

18 *

Cloud chamber

○ Yes ○ No

19 *

Experiments with the electron tube

○ Yes ○ No

20 *

Experiments with the X-ray machine

○ Yes ○ No

21 *

Positron-Emission-Tomography

○ Yes ○ No

22 *

Superconductivity

○ Yes ○ No

23 *

Paul-Trap

○ Yes ○ No

How did you like the experiments in S'Cool LAB?

Page 6

24 * **Cloud Chamber**

not at all ○ ○ ○ ○ ○ ○ ○ very much

Page 7

25 * **Experiments with the electron tube**

not at all ○ ○ ○ ○ ○ ○ ○ very much

Page 8

26 * **Experiments with the X-ray machine**

not at all ○ ○ ○ ○ ○ ○ ○ very much

Page 12

Mental stress during the time in S'Cool LAB

30 * **How high was the level of mental stress** you experienced in S'Cool LAB?

very low ○ ○ ○ ○ ○ ○ ○ very high

31 * **How hard did you have to work** to perform the experiments?

very little ○ ○ ○ ○ ○ ○ ○ very much

32 * How **frustrated** were you during your time in S'Cool LAB?

very little ○ ○ ○ ○ ○ ○ ○ very much

33 * How do you rate the **duration** of the different experiments?

too short ○ ○ ○ ○ ○ ○ ○ too long

34 * How do you rate the **number** of different experiments you were working on?

not enough ○ ○ ○ ○ ○ ○ ○ too many

Page 13

35 *

Information about your working group in S'Cool LAB

Including yourself, who was working at your table?

	0	1	2	3	4	5	6
How many students were in your working group?	○	○	○	○	○	○	○
How many members of your working group were female?	○	○	○	○	○	○	○
How many members of your working group were male?	○	○	○	○	○	○	○

Page 14

36 *

How much did you enjoy the time in S'Cool LAB?

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
I enjoyed using the equipment.	○	○	○	○	○	○
I enjoyed preparing the experiments.	○	○	○	○	○	○
I enjoyed doing the experiments.	○	○	○	○	○	○
I enjoyed discussing the phenomena.	○	○	○	○	○	○
I enjoyed the time in S'Cool LAB.	○	○	○	○	○	○
I will recommend S'Cool LAB to my friends.	○	○	○	○	○	○
I liked the look of S'Cool LAB.	○	○	○	○	○	○

37 * **How easy was the work in S'Cool LAB for you?**

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
It was easy to understand the instructions.	○	○	○	○	○	○
It was easy to prepare the experiments.	○	○	○	○	○	○
It was easy to use the equipment.	○	○	○	○	○	○
It was easy to get a result.	○	○	○	○	○	○
It was easy to understand what it was all about.	○	○	○	○	○	○

38 * **Was your time in S'Cool LAB helpful for you?**

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
The time in S'Cool LAB was helpful for my physics classes.	○	○	○	○	○	○
The time in S'Cool LAB helped me to get better at doing experiments.	○	○	○	○	○	○
The time in S'Cool LAB helped me to understand physics concepts.	○	○	○	○	○	○
The time in S'Cool LAB helped me to get an idea about how scientists work.	○	○	○	○	○	○
The time in S'Cool LAB helped me to get an idea about what CERN does.	○	○	○	○	○	○

39 * Organisation and way of working in S'Cool LAB

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
I was familiar with the schedule in S'Cool LAB.	○	○	○	○	○	○
The time in S'Cool LAB was well organized.	○	○	○	○	○	○
I was well prepared for my time in S'Cool LAB.	○	○	○	○	○	○
I knew what to expect in S'Cool LAB.	○	○	○	○	○	○
I was easily getting familiar with the different work places in S'Cool LAB.	○	○	○	○	○	○
Before arriving at CERN, I knew how I was going to work in S'Cool LAB.	○	○	○	○	○	○
The materials needed were easy to find.	○	○	○	○	○	○
It was hard for me to do the experiments.	○	○	○	○	○	○
I was afraid to make mistakes when doing the experiments.	○	○	○	○	○	○
It took a long time until I was familiar with the equipment and was able to start the experiments.	○	○	○	○	○	○
It was fun to try out the experiments.	○	○	○	○	○	○
I received enough information to perform the experiments.	○	○	○	○	○	○
I was able to concentrate on the experiments, without "fighting" with the equipment.	○	○	○	○	○	○
The instructions were helpful for doing the experiments.	○	○	○	○	○	○

Page 16

40 * Did you feel “cognitively activated” in S’Cool LAB?

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
I was focused when doing the experiments in S'Cool LAB.	○	○	○	○	○	○
I was actively involved in doing the experiments in S'Cool LAB.	○	○	○	○	○	○
I took a critical look at new ideas or material.	○	○	○	○	○	○
I tried to distinguish between important and unimportant things when doing the experiments.	○	○	○	○	○	○
During the experiments I tried to connect what I was learning with my previous physics knowledge.	○	○	○	○	○	○

Page 17

41 * **Did you feel overwhelmed in S'Cool LAB?**

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
I had problems interpreting my observations correctly.	○	○	○	○	○	○
The experiments were difficult.	○	○	○	○	○	○
I had problems understanding the physics principles of the experiments.	○	○	○	○	○	○
It was hard for me to understand the instructions.	○	○	○	○	○	○
I would have needed more assistance when working with the experiments.	○	○	○	○	○	○

Page 18

How mentally demanding was the use of the equipment?

Page 19

42 * **Cloud Chamber**

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
It was difficult to use the equipment.	○	○	○	○	○	○
I was overwhelmed by the features provided by the equipment.	○	○	○	○	○	○
I had problems with using the equipment to make observations.	○	○	○	○	○	○
I had problems understanding the physics principles of this experiment.	○	○	○	○	○	○

Page 20

43 * Experiments with the electron tube

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
It was difficult to use the equipment.	○	○	○	○	○	○
I was overwhelmed by the features provided by the equipment.	○	○	○	○	○	○
I had problems with using the equipment to make observations.	○	○	○	○	○	○
I had problems understanding the physics principles of this experiment.	○	○	○	○	○	○

Page 21

44 * Experiments with the X-ray machine

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
It was difficult to use the equipment.	○	○	○	○	○	○
I was overwhelmed by the features provided by the equipment.	○	○	○	○	○	○
I had problems with using the equipment to make observations.	○	○	○	○	○	○
I had problems understanding the physics principles of this experiment.	○	○	○	○	○	○

Page 22

Page 25

48 * What do you think about the tutors in S'Cool LAB?

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
I had the chance to ask the tutors in S'Cool LAB questions.	○	○	○	○	○	○
I had the feeling that the tutors are fascinated by physics.	○	○	○	○	○	○
The discussions with the tutors were important to me.	○	○	○	○	○	○
I liked the working atmosphere in S'Cool LAB.	○	○	○	○	○	○
The tutors responded to my questions and comments while doing experiments.	○	○	○	○	○	○
The tutors helped me with problems while doing the experiments.	○	○	○	○	○	○

49 **Would you like to leave a message to the tutors in S'Cool LAB?**

Paragraph

Path: p

Page 26

50 * **How did you like your time in S'Cool LAB?**

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
After my visit in S'Cool LAB I spent time on particle physics in my free time – even beyond homework.	○	○	○	○	○	○
During my time in S'Cool LAB, I enjoyed solving physics problems.	○	○	○	○	○	○
I discussed what I have learned in S'Cool LAB with my friends or my family.	○	○	○	○	○	○
When I was working on physics problems in S'Cool LAB, I did not notice time passing.	○	○	○	○	○	○
I liked my time in S'Cool LAB.	○	○	○	○	○	○
I would like to spend more time in S'Cool LAB.	○	○	○	○	○	○

Page 27

51 * What do you think about your performance during your time in S'Cool LAB?

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
I understood the learning material in S'Cool LAB.	○	○	○	○	○	○
I looked forward to my time in S'Cool LAB.	○	○	○	○	○	○
My performance in S'Cool LAB was good.	○	○	○	○	○	○
I found it easy to solve tasks in S'Cool LAB.	○	○	○	○	○	○
I believe that during my time in S'Cool LAB my classmates thought I am good in physics.	○	○	○	○	○	○
Doing experiments in S'Cool LAB was easy for me.	○	○	○	○	○	○
I was good at doing experiments in S'Cool LAB.	○	○	○	○	○	○
During my time in S'Cool LAB, I found it easy to finish in the time given.	○	○	○	○	○	○
The experiments in S'Cool LAB helped me to understand the learning material.	○	○	○	○	○	○

Page 28

52 * Are you curious about particle physics?

Please rate the following statements.

	completely disagree	disagree	disagree a little	agree a little	agree	agree completely
My time in S'Cool LAB drew my attention to the topic of particle physics.	○	○	○	○	○	○
I want to learn more about particle physics.	○	○	○	○	○	○
The time in S'Cool LAB made me curious to learn more about particle physics.	○	○	○	○	○	○
I am more curious about particle physics now, after participating in S'Cool LAB activities.	○	○	○	○	○	○
I want to understand the topic of particle physics better.	○	○	○	○	○	○
I find it fascinating to spend time on particle physics.	○	○	○	○	○	○

Page 29

53 * **How important were the following aspects of your trip to CERN for you?**

	not at all important	not important	rather not important	rather important	important	very important
Doing experiments on my own	○	○	○	○	○	○
Having fun	○	○	○	○	○	○
No school	○	○	○	○	○	○
Learning more about physics	○	○	○	○	○	○
Learning more about particle physics	○	○	○	○	○	○
Having visited S'Cool LAB	○	○	○	○	○	○
Having visited CERN	○	○	○	○	○	○

Page 30

54 * **Language problems**

Did you have to deal with foreign languages in S'Cool LAB?

○ Yes ○ No

Page 31

55 * How difficult did the foreign languages make it for you to understand the instructions in S'Cool LAB?

not at all difficult ○ ○ ○ ○ ○ ○ ○ very difficult

56 * How difficult did the foreign languages make it for you to understand the the physics principles in S'Cool LAB?

not at all difficult ○ ○ ○ ○ ○ ○ ○ very difficult

Page 32

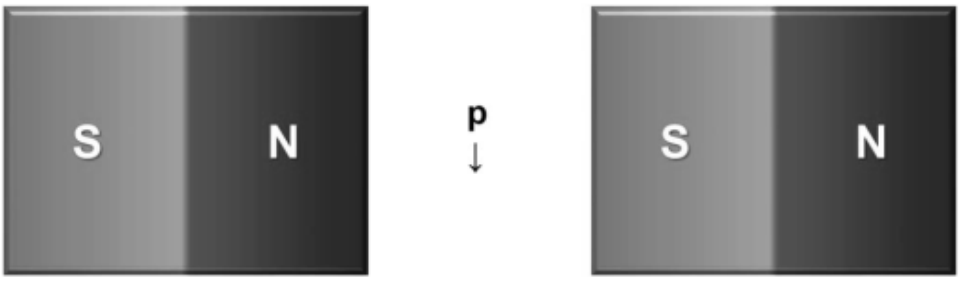
Some questions about physics

You will not know the answer to all of the following questions yet. However, it is extremely important that you fill in the questionnaire ***without help from your teacher or your classmates*** *because we are interested in your ideas!*

If you are not sure about one of the questions, just guess and select the answer option which corresponds best to your opinion. ***This is not a test****! You will not be judged for your answers in any way.*

Page 33

57 * **A proton is moving downwards at constant speed through the uniform magnetic field between two bar magnets.**



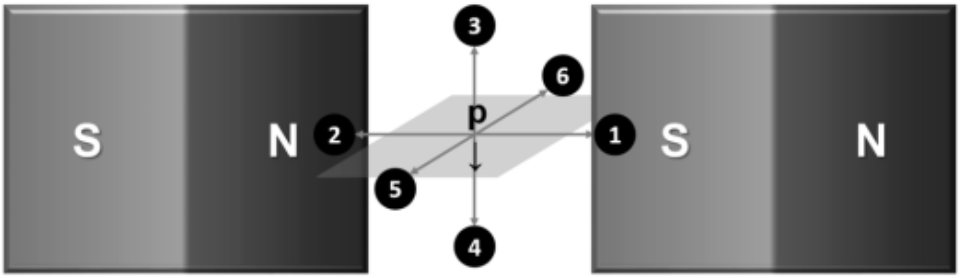
Do the magnets exert a force on the proton?

- ○ Yes, the magnets exert a force on the proton.
- ○ No, the magnets do not exert a force on the proton.

Page 34

58 * **If the magnets exert a force on the proton: In which direction does the force point?**

The coordinate system below might help you to identify the direction.



- ○ to the right (1)
- ○ to the left (2)
- ○ upwards (3)
- ○ downwards (4)
- ○ out of the page (5)
- ○ into the page (6)
- ○ in a different direction

Page 35

59 * **Explanation:**

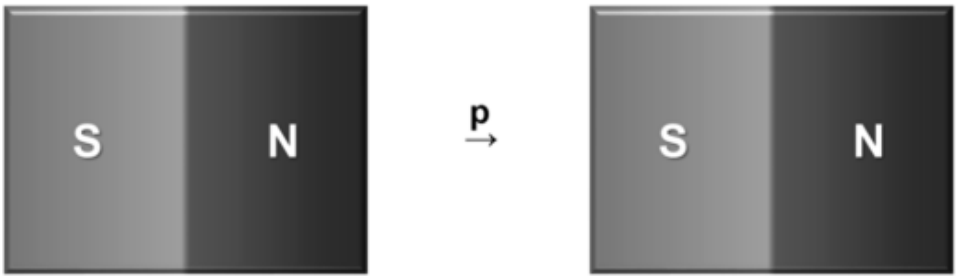
- ○ The south pole S of the magnets attracts the proton.
- ○ The north pole N of the magnets attracts the proton.
- ○ The magnetic force acts in the direction of the magnetic field.
- ○ The magnetic force acts in the direction of the movement of the proton.
- ○ The magnetic force acts on the proton according to the right-hand-rule.
- ○ None of the explanations above apply.

60 * **How confident are you that your answers are correct?**

not confident at all ○ ○ ○ ○ ○ very confident

Page 36

61 * **A proton is moving to the right at constant speed through the uniform magnetic field between two bar magnets.**



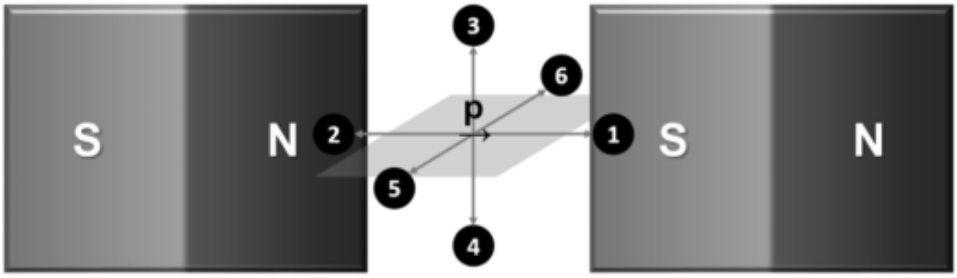
Do the magnets exert a force on the proton?

- ○ Yes, the magnets exert a force on the proton.
- ○ No, the magnets do not exert a force on the proton.

Page 37

62 * **If the magnets exert a force on the proton: In which direction does the force point?**

The coordinate system below might help you to identify the direction.



- ○ to the right (1)
- ○ to the left (2)
- ○ upwards (3)
- ○ downwards (4)
- ○ out of the page (5)
- ○ into the page (6)
- ○ in a different direction

Page 38

63 * **Explanation:**

- ○ The south pole S of the magnets attracts the proton.
- ○ The north pole N of the magnets attracts the proton.
- ○ The magnetic force acts in the direction of the magnetic field.
- ○ The magnetic force acts in the direction of the movement of the proton.
- ○ The magnetic force acts on the proton according to the right-hand-rule.
- ○ None of the explanations above apply.

64 * **How confident are you that your answers are correct?**

not confident at all ○ ○ ○ ○ ○ very confident

Page 39

65 * **A proton is at rest in the uniform magnetic field between two bar magnets.**

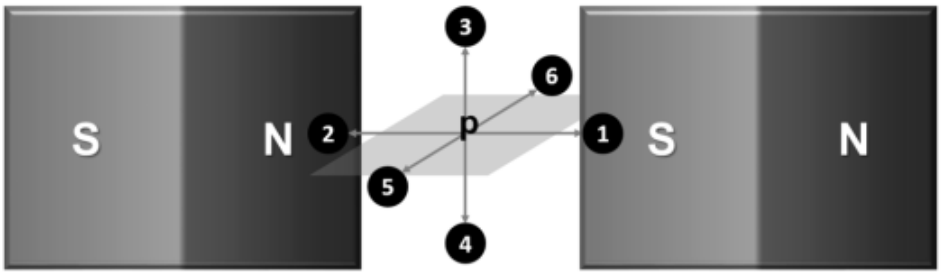
Do the magnets exert a force on the proton?

- ○ Yes, the magnets exert a force on the proton.
- ○ No, the magnets do not exert a force on the proton.

Page 40

66 * **If the magnets exert a force on the proton: In which direction does the force point?**

The coordinate system below might help you to identify the direction.



- ○ to the right (1)
- ○ to the left (2)
- ○ upwards (3)
- ○ downwards (4)
- ○ out of the page (5)
- ○ into the page (6)
- ○ in a different direction

Page 41

67 * **Explanation:**

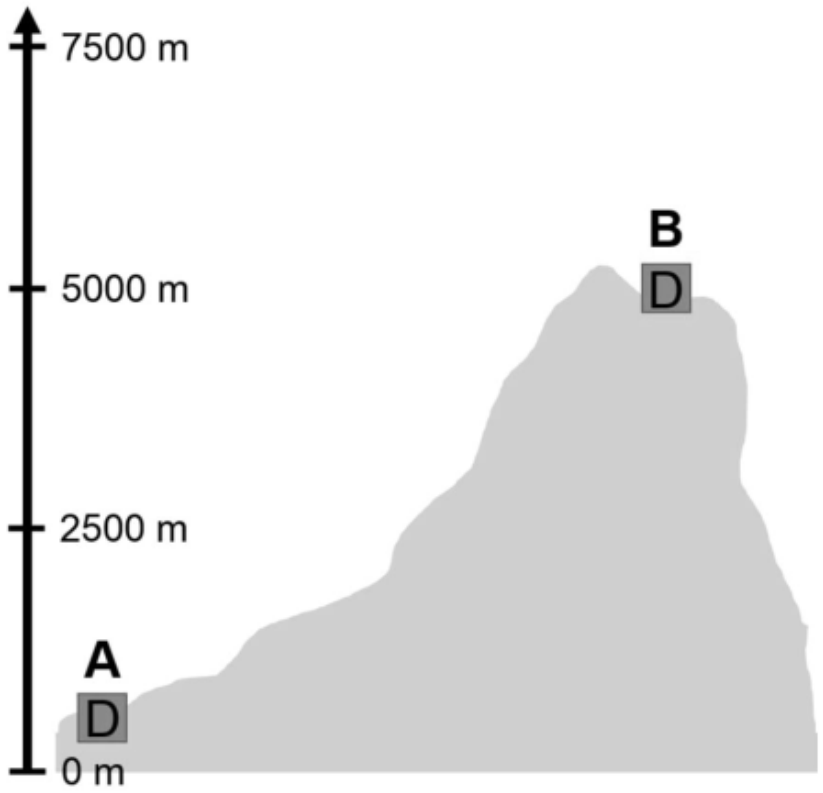
- ○ The south pole S of the magnets attracts the proton.
- ○ The north pole N of the magnets attracts the proton.
- ○ The magnetic force acts in the direction of the magnetic field.
- ○ The magnetic force acts in the direction of the movement of the proton.
- ○ The magnetic force acts on the proton according to the right-hand-rule.
- ○ None of the explanations above apply.

68 * **How confident are you that your answers are correct?**

not confident at all ○ ○ ○ ○ ○ very confident

Page 42

69 * **Imagine you compare two particle detectors [D]. One is placed in the valley at 0 m altitude. The other one is placed on top of a mountain at 5000 m altitude.**



Will the particle detector on top of the mountain detect a higher number of particles than the detector at 0 m?

- ○ Yes, the particle detector on top of the mountain will detect a HIGHER number of particles.
- ○ No, the particle detector on top of the mountain will detect a LOWER number of particles.
- ○ No, there won't be a difference.

70 * **Explanation:**

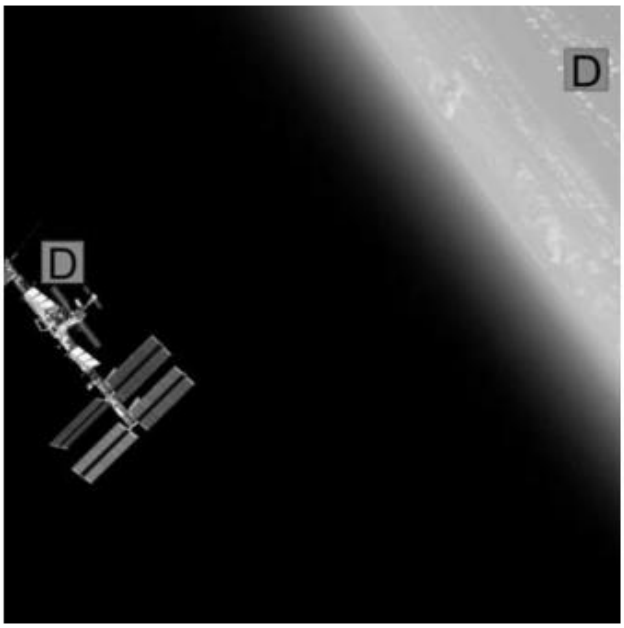
- ○ The properties of the atmosphere on top of the mountain and in the valley are the same.
- ○ The properties of the ground on top of the mountain and in the valley are the same.
- ○ The Earth is the main source of the detected particles.
- ○ Space is the main source of the detected particles.
- ○ The atmosphere has a shielding effect.
- ○ There are fewer air molecules on top of the mountain.
- ○ different explanation:

71 * **How confident are you that your answers are correct?**

not confident at all ○ ○ ○ ○ ○ very confident

Page 43

72 * **Imagine you compare two particle detectors [D]. One is placed on the surface of the Earth at 0 km altitude. The other one is placed on a satellite at 400 km altitude.**



Will the particle detector on the satellite detect a higher number of particles than the detector on the surface of the Earth?

- ○ Yes, the particle detector on the satellite will detect a HIGHER number of particles.
- ○ No, the particle detector on the satellite will detect a LOWER number of particles.
- ○ No, there won't be a difference.

73 * **Explanation:**

- ○ The properties of the atmosphere at 400 km altitude and at 0 km altitude are the same.
- ○ The Earth is the main source of the detected particles.
- ○ Space is the main source of the detected particles.
- ○ The atmosphere has a shielding effect.
- ○ There are fewer air molecules in space.
- ○ different explanation:

74 * **How confident are you that your answers are correct?**

not confident at all ○ ○ ○ ○ ○ very confident

Page 44

75 * **Is it dangerous for the cleaning staff of a hospital when they clean a room with an X-ray machine?**

- ○ Yes, parts of the X-ray device are exposed to X-rays and therefore become radioactive themselves.
- ○ Yes, because X-ray machines are always a source of radiation, even if they are switched off.
- ○ No, there is no danger, because the X-ray units are switched off during the cleaning.
- ○ No, provided that the entire air of the treatment rooms has been exchanged before the cleaning.
- ○ No, provided that the cleaning staff waits a while before they enter the room.

76 * **How confident are you that your answer is correct?**

not confident at all ○ ○ ○ ○ ○ very confident

Page 45

77 * **A watch is X-rayed for 1 hour inside an X-ray machine. Is it dangerous to wear the watch afterwards?**

- ○ Yes, ionizing radiation accumulates in the watch, therefore the watch will emit radiation afterwards.
- ○ Yes, the watch has been exposed to X-rays and therefore the atoms in the watch become radioactive.
- ○ No, to become radioactive itself, the watch needs to be exposed to X-rays for a longer time.
- ○ No, as long as no radioactive atoms are transferred to the watch, the watch does not emit radiation.

78 * **How confident are you that your answer is correct?**

not confident at all ○ ○ ○ ○ ○ very confident

Page 46

79 * **Feedback about the questionnaire**

What do you think about the length of this questionnaire?

very short ○ ○ ○ ○ ○ ○ ○ very long

80 * How difficult was it for you to understand the questions in this questionnaire?

very easy ○ ○ ○ ○ ○ ○ ○ very difficult

81 **Do you have comments about specific questions or the questionnaire in general?**

Paragraph

Path: p

82 **Is there anything else you wanted to tell us?**

Paragraph

Path: p

You did it! Thank you very much for your help!

Please don't forget to submit your questionnaire by clicking the button "submit the questionnaire" below.

10.6 Overview of sample of the pilot study

Table 135. Description of the sample of the pilot study with the number of completed pre-visit and post-visit questionnaires as well as their mode of administration.

Country	Grade	Date of CERN visit	# pre	mode	# post	mode
Germany	11	11/05/2016	23	on-site	6	online
Turkey	9	18/05/2016	36	on-site	24	online
Sweden	12	25/05/2016	29	on-site	12	online
Germany	11	03/06/2016	18	on-site	18	on-site[50]
		Sum:	**106**	**Sum:**	**60**	

[50] These questionnaires were excluded from some of the scale analyses because some of the items are not sensitive at a direct evaluation after the hands-on session in S'Cool LAB, for example, *"I spent time on physics I learned about in S'Cool LAB in my free time – even beyond homework."*

10.7 Main study: additional item and instrument properties

10.7.1 Clustering of students: Countries, teachers, groups

Student group

#	% students
1	1.4%
2	1.6%
3	6.3%
4	3.3%
5	2.9%
6	5.3%
7	4.5%
8	4.9%
9	2.8%
10	4.7%
11	2.4%
12	3.5%
13	4.3%
14	6.5%
15	3.9%
16	4.1%
17	2.2%
18	4.3%
19	3.1%
20	4.5%
21	4.1%
22	2.4%
23	5.7%
24	1.6%
25	4.7%
26	2.2%
27	1.8%
28	1.0%

Teacher

#	% students
1	1.0%
2	0.8%
5	1.0%
6	2.2%
7	4.3%
8	0.4%
9	1.4%
10	0.2%
11	5.7%
12	2.4%
13	1.6%
14	1.0%
15	1.6%
16	4.5%
17	3.1%
18	4.3%
19	0.2%
20	2.2%
21	3.9%
25	2.0%
26	2.0%
27	2.6%
28	1.2%
29	1.4%
30	0.2%
31	0.2%
32	0.4%
33	0.6%
34	2.0%
35	0.6%

#	% students
36	1.4%
37	0.4%
38	3.3%
39	0.2%
40	2.0%
41	0.4%
43	2.2%
44	2.6%
45	2.8%
46	4.9%
50	4.5%
51	2.4%
52	2.9%
53	1.2%
54	0.6%
55	0.6%
56	0.2%
57	0.4%
59	3.3%
60	2.2%
61	0.2%
62	2.2%
63	1.8%
64	1.0%
65	0.4%
66	0.2%
67	0.4%
68	0.4%
69	0.6%

School country

#	% students
AT	5%
CZ	3%
FR	16%
DE	3%
IT	7%
NL	2%
NO	9%
SG	4%
PL	6%
PO	4%
ES	15%
CH	8%
UK	19%

10.7.2 Gender (SEX)

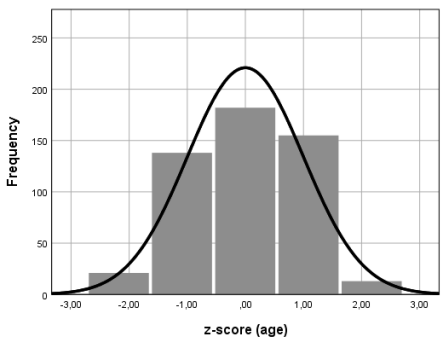
	% students
Female	36%
Male	64%
N	509
missing	0

10.7.3 Age versus grade (AGE)

Age of the participants main study		**School grade as reported by teachers**	
age in years	**% students**	**grade**	**% students**
15	4%	**10**	5%
16	27%	**11**	22%
17	36%	**12**	41%
18	30%	**13**	32%
19	3%		
		N	509
N	509	**missing**	0
missing	0	**M**	12
M	17.0	**SD**	0.9
SD	0.9		

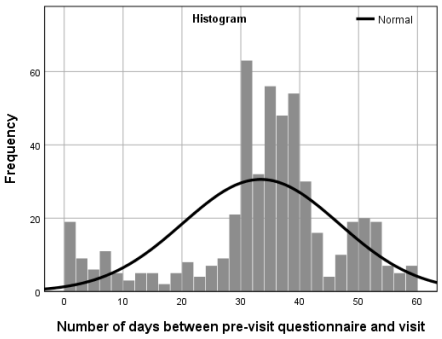
Because of difficulty to compare school grades between different school systems, it was decided to drop the variables grade and use age instead as control variable.

z-score: age (Z_AGE)	
Skewness	-0.13
SE	0.11
z-value	1 < 3.29, n.s.
Kurtosis	-0.68
SE	0.22
z-value	3.1 < 3.29, n.s.
Kolmo	KS(509)= 0.19, p < 0.001
Shapiro	SW(509) = 0.89, p < 0.001

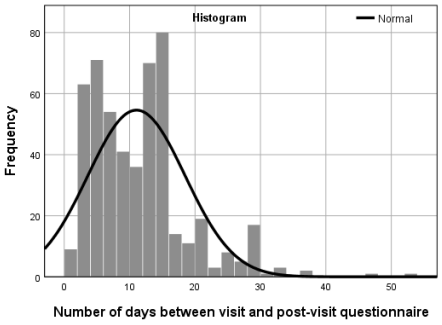


10.7.4 When did students fill out the online questionnaires?

Number of days between pre-visit questionnaire and visit	
Mean (SD)	**33 days (13 days)**
Skewness	**-0.80**
SE	0.11
z-value	7 > 3.29, significant
Kurtosis	**0.54**
SE	0.22
z-value	2.5 < 3.29, n.s.
Kolmo	KS(509)= 0.17, p < 0.001
Shapiro	SW(509) = 0.92, p < 0.001



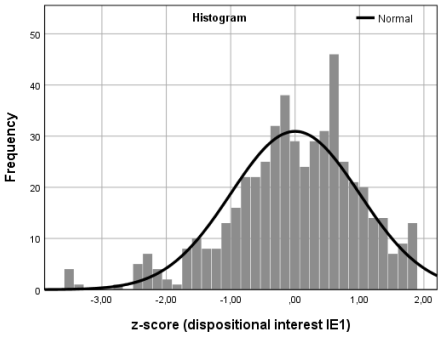
Number of days between visit and post-visit questionnaire	
Mean (SD)	**11 days (7 days)**
Skewness	**1.33**
SE	0.11
z-value	12 > 3.29, significant
Kurtosis	**2.99**
SE	0.22
z-value	14 >> 3.29, significant
Kolmo	KS(509)= 0.13, p < 0.001
Shapiro	SW(509) = 0.90, p < 0.001



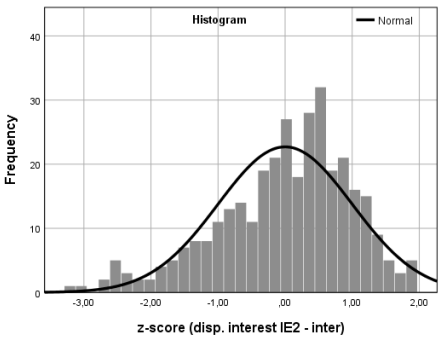
10.7.5 Dispositional and situational interest in physics (IE1, IE2, IE3)

Extreme cases	Extreme values	Handling
133	z(IE1)=-3.4	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
146	z(IE1)=-3.4, z(IE3)=-3.8	
196	z(IE1)=-3.4	
326	z(IE1)=-3.4, z(IE3)=-4.4	
327	z(IE1)=-3.4	
383	Z(IE3)=-3.8	

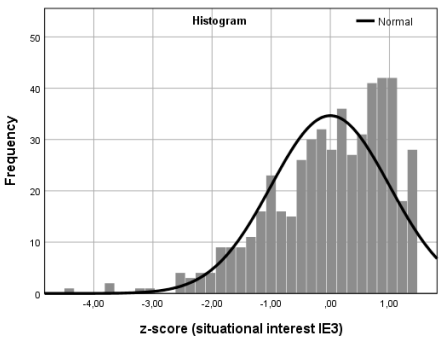
z-score: dispositional interest physics (Z_IE1)	
Skewness	**-0.66**
SE	0.11
z-value	6 > 3.29, significant
Kurtosis	**0.72**
SE	0.22
z-value	3.27 < 3.29, n.s.
Kolmo	KS(509)= 0.07, p < 0.001
Shapiro	SW(509) = 0.97, p < 0.001



z-score: situational interest (Z_IE2) n = 335	
Skewness	**-0.68**
SE	0.12
z-value	6 > 3.29, significant
Kurtosis	**0.26**
SE	0.27
z-value	1 < 3.29, n.s.
Kolmo	KS(335)= 0.10 p < 0.001
Shapiro	SW(335) = 0.97, p < 0.001



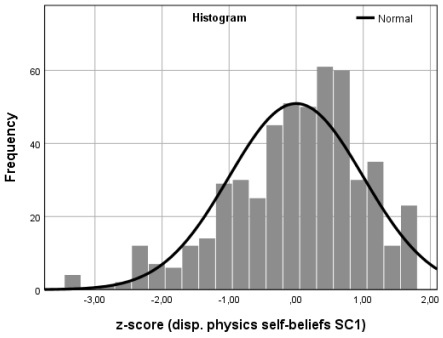
z-score: situational interest (Z_IE3)	
Skewness	**-0.90**
SE	0.11
z-value	8 > 3.29, significant
Kurtosis	**0.85**
SE	0.22
z-value	4 > 3.29, n.s.
Kolmo	KS(509)= 0.10 p < 0.001
Shapiro	SW(509) = 0.94, p < 0.001



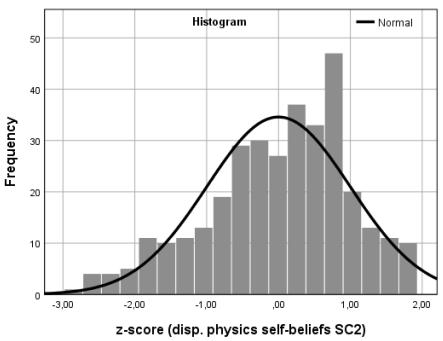
10.7.6 Dispositional and situational physics self-beliefs (SC1, SC2, SC3)

Extreme cases	Extreme values	Handling
137	z(SC1)=-3.3	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
146	z(SC1)=-3.3	
196	z(SC1)=-3.3	
326	z(SC1)=-3.3	
434	Z(SC3)=-4.8	
187	Z(SC3)=-4.1	

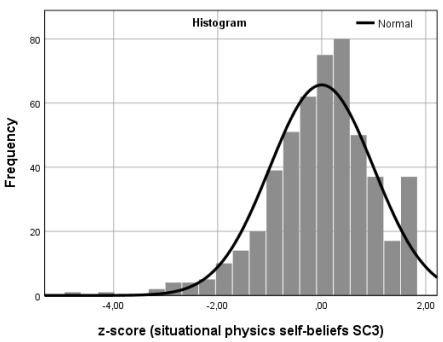
z-score: dispositional self-beliefs physics (Z_SC1)	
Skewness	**-0.68**
SE	0.11
z-value	6 > 3.29, significant
Kurtosis	**0.39**
SE	0.22
z-value	1.8 < 3.29, n.s.
Kolmo	KS(509)= 0.10, p < 0.001
Shapiro	SW(509) = 0.96, p < 0.001



z-score: dispositional self-beliefs physics (Z_SC2)	
Skewness	**-0.51**
SE	0.13
z-value	4 > 3.29, significant
Kurtosis	**-0.16**
SE	0.27
z-value	0.6 < 3.29, n. s.
Kolmo	KS(335)= 0.11 p < 0.001
Shapiro	SW(335) = 0.97, p < 0.001



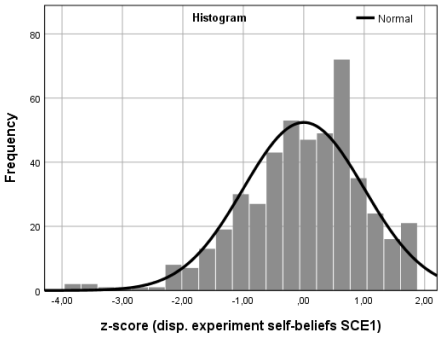
z-score: situational self-beliefs physics (Z_SC3)	
Skewness	**-0.73**
SE	0.11
z-value	7 > 3.29, significant
Kurtosis	**1.48**
SE	0.22
z-value	7 > 3.29, significant
Kolmo	KS(509)= 0.11, p < 0.001
Shapiro	SW(509) = 0.96, p < 0.001



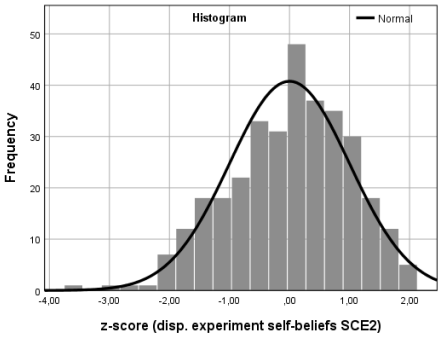
10.7.7 Disp. and situat. experiments self-beliefs (SCE1, SCE3, SCE3)

Extreme cases	Extreme values	Handling
137	z(SCE1)=-3.5	Excluded from further analysis in favour of physics self-beliefs due to smaller sample size and high number of extreme values
146	z(SCE1)=-3.8, z(SCE2)=-3.6	
191	z(SCE1)=-3.5	
326	z(SCE1)=-3.8	
434	z(SCE3)=-4.8	
187	z(SCE3)=-4.8	
320	z(SCE3)=-3.5	
294	z(SCE3)=-3.5	

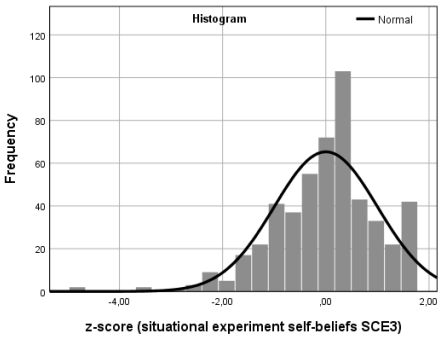
z-score: dispositional experiment self-beliefs (Z_SCE1)	
Skewness	**-0.69**
SE	0.11
z-value	6 > 3.29, significant
Kurtosis	**0.90**
SE	0.22
z-value	4 > 3.29, significant
Kolmo	KS(472)= 0.10, p < 0.001
Shapiro	SW(472) = 0.96, p < 0.001



z-score: dispositional experiment self-beliefs (Z_SCE2)	
Skewness	**-0.40**
SE	0.13
z-value	3.1 < 3.29, n. s.
Kurtosis	**-0.07**
SE	0.27
z-value	0.3 < 3.29, n. s.
Kolmo	KS(330)= 0.11 p < 0.001
Shapiro	SW(330) = 0.98, p < 0.001



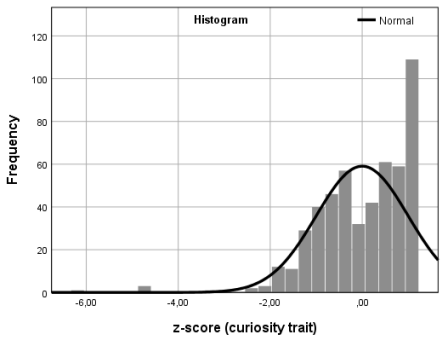
z-score: situational experiment self-beliefs (Z_SCE3)	
Skewness	**-0.81**
SE	0.11
z-value	7 > 3.29, significant
Kurtosis	**1.92**
SE	0.22
z-value	9 > 3.29, significant
Kolmo	KS(509)= 0.12, p < 0.001
Shapiro	SW(509) = 0.95, p < 0.001



10.7.8 Curiosity trait (CT)

Extreme cases	Extreme values	Handling
269	z(CT1)=-6.2	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
353	z(CT1)=-4.7	
380	z(CT1)=-4.7	
398	z(CT1)=-4.7	

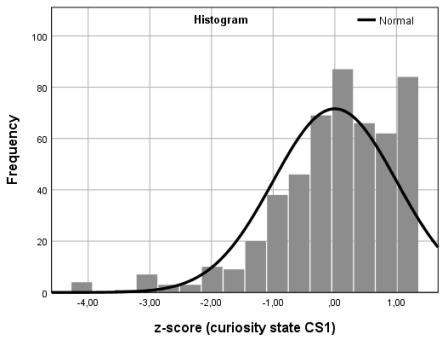
z-score: curiosity trait (Z_CT)	
Skewness	**-1.41**
SE	0.11
z-value	13 >> 3.29, significant
Kurtosis	**4.51**
SE	0.22
z-value	22 >> 3.29, significant
Kolmo	KS(509)= 0.14, p < 0.001
Shapiro	SW(509) = 0.87, p < 0.001



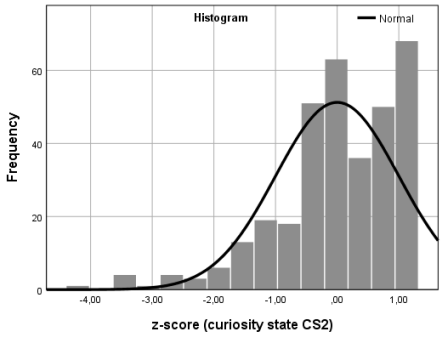
10.7.9 Curiosity state particle physics (CS1, CS2, CS3)

Extreme cases	Extreme values	Handling
157	z(CS1)=-4.1	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
196	z(CS1)=-3.4	
269	z(CS1)=-4.1	
326	z(CS1)=-4.1, z(CS2)=-3.4, z(CS3)=-3.9	
353	z(CS1)=-4.1	
386	z(CS3)=-3.9	
146	z(CS2)=-3.4, z(CS3)=-3.4	
327	z(CS2)=-4.2	
323	z(CS2)=-3.4	
305	z(CS2)=-3.4	

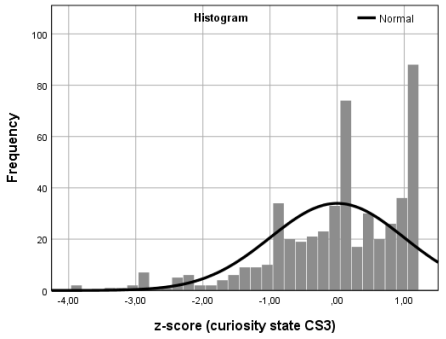
z-score: curiosity state particle physics (Z_CS1)	
Skewness	**-1.24**
SE	0.11
z-value	9 > 3.29, significant
Kurtosis	**2.21**
SE	0.22
z-value	10 > 3.29, significant
Kolmo	KS(509)= 0.14, p < 0.001
Shapiro	SW(509) = 0.90, p < 0.001



z-score: curiosity state particle physics (Z_CS2)	
Skewness	**-1.17**
SE	0.11
z-value	11 >> 3.29, significant
Kurtosis	**2.08**
SE	0.22
z-value	9 > 3.29, significant
Kolmo	KS(509)= 0.14, p < 0.001
Shapiro	SW(509) = 0.90, p < 0.001



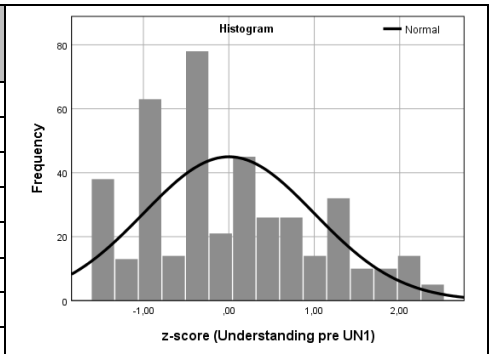
z-score: curiosity state particle physics (Z_CS3)	
Skewness	**-1.10**
SE	0.11
z-value	10 >> 3.29, significant
Kurtosis	**1.28**
SE	0.22
z-value	6 > 3.29, significant
Kolmo	KS(509)= 0.15, p < 0.001
Shapiro	SW(509) = 0.88, p < 0.001



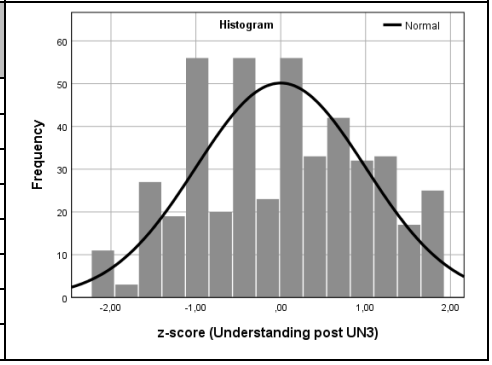
10.7.10 Conceptual understanding (UN1 & UN3)

No extreme cases identified.

z-score: Understanding pre (Z_UN1)	
Skewness	**0.48**
SE	0.12
z-value	4 > 3.29, significant
Kurtosis	**-0.59**
SE	0.24
z-value	2.5 < 3.29, n. s.
Kolmo	KS(409)= 0.15, p < 0.001
Shapiro	SW(409) = 0.95, p < 0.001



z-score: Understanding post (Z_UN3)	
Skewness	**-0.03**
SE	0.12
z-value	0 < 3.29, n. s.
Kurtosis	**-0.86**
SE	0.23
z-value	3.7 > 3.29, significant
Kolmo	KS(453)= 0.09, p < 0.001
Shapiro	SW(453) = 0.97, p < 0.001



10.7.10.1 Grading of items related to the electron tube workshop PD, PR, PS

			pre		post		
Code	**Formulae**	**Description**	**M**	**SD**	**M**	**SD**	
pd_ 1	1: AND(pd_yn=1; pd_dir=5; pd_expl=5)	Correct	0.16	0.37	0.29	0.45	discarded
pd_2	1: AND(pd_yn=1; OR(pd_dir=5;pd_dir=6); pd_expl=5)	Easier	0.21	0.41	0.43	0.50	discarded
pd_3	0.5: AND(pd_yn=1; pd_dir=5) (pd_expl wrong) 1: AND(pd_yn=1; pd_dir=5; pd_expl=5)	Stepwise 2 level	0.25	0.38	0.36	0.44	discarded
pd_4	0.5: AND(pd_yn=1; OR(pd_dir=5; pd_dir=6)) 1: AND(pd_yn=1; OR(pd_dir=5; pd_dir=6); pd_expl=5)	Stepwise easier, final	0.28	0.38	0.45	0.41	chosen
pr_1	1: AND(pd_yn=2; OR(pd_expl=5; pd_expl=6)	Correct	0.16	0.37	0.29	0.45	chosen
ps_1	1: AND(pd_yn=2; OR(pd_expl=5; pd_expl=6)	Correct	0.28	0.45	0.31	0.46	chosen

10.7.10.2 PD: A proton is moving downwards between two bar magnets

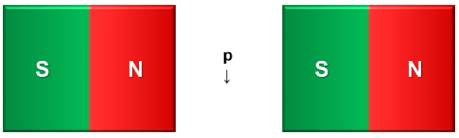
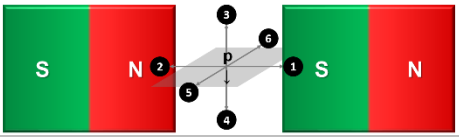
code	Text and picture	Answer options
pd_yn	**A proton is moving downwards at constant speed through the uniform magnetic field between two bar magnets.** **Do the magnets exert a force on the proton?**  1=Yes, the magnets exert a force on the proton. 2=No, the magnets do not exert a force on the proton.	
pd_dir	**If the magnets exert a force on the proton: In which direction does the force point?** The coordinate system below might help you to identify the direction.  	1=to the right (1). 2=to the left (2). 3=upwards (3). 4=downwards (4). 5=out of the page (5). 6=into the page (6). 7=in a different direction.
pd_expl	**Explanation:** 1=The south pole S of the magnets attracts the proton. 2=The north pole N of the magnets attracts the proton. 3=The magnetic force acts in the direction of the magnetic field. 4=The magnetic force acts in the direction of the movement of the proton. 5=The magnetic force acts on the proton according to the right-hand-rule. 6=None of the explanations above apply.	
pd_sure	**How confident are you that your answers are correct?** 1= not confident at all|very confident =5	

Table 136. Item PD: Students' answers in the pre- and post-visit questionnaire.

	N_{pre}	%	N_{post}	%
Tier-1: Do the magnets exert a force on the proton?	**509**		**509**	
Yes	461	91%	467	92%
No	48	9%	42	8%
Tier-2: If yes, which direction?	**461**		**453**	
1=to the right (1)	98	19%	61	12%
2=to the left (2)	63	12%	33	6%
3=upwards (3)	20	4%	29	6%
4=downwards (4)	48	9%	43	8%
5=out of the page (5)	180	35%	196	39%
6=into the page (6)	29	6%	84	17%
7=in a different direction	23	5%	7	1%

	N_{pre}	%	N_{post}	%
Tier-3: Explanation	**492**		**509**	
1=The south pole S of the magnets attracts the proton.	64	13%	47	9%
2=The north pole N of the magnets attracts the proton.	56	11%	36	7%
3=The magnetic force acts in the direction of the magnetic field.	93	18%	76	15%
4=The magnetic force acts in the direction of the movement of the proton.	39	8%	34	7%
5=The magnetic force acts on the proton according to the right-hand-rule.	151	30%	258	51%
6=None of the explanations above apply.	89	17%	58	11%
Tier-4: How confident are you that your answers are correct?	**492**		**509**	
1= not confident at all	90	18%	66	13%
2	130	26%	104	20%
3	135	27%	137	27%
4	88	17%	119	23%
5=very confident	49	10%	83	16%
Average confidence		**44%**		**53%**

Table 137. Item PD: Combinations of tier-2 and tier-3 level answer choices during the pre- and post-visit questionnaire. Combination frequencies larger than two times the guessing ratio are highlighted in orange. The correct combination is marked in blue, answers that were also rated correct are marked in green.

PRE		Tier-3 explanation pre						
		1	**2**	**3**	**4**	**5**	**6**	SUM
Tier-2 direction pre	**1**	10.4%	0.5%	7.4%	0.2%	2.9%	0.5%	21.8%
	2	0.9%	8.6%	2.3%	0.5%	1.8%	0.0%	14.0%
	3	0.5%	0.5%	0.7%	0.5%	1.6%	0.9%	4.5%
	4	0.9%	0.5%	3.6%	2.9%	2.5%	0.5%	10.8%
	5	0.7%	1.8%	1.8%	1.6%	**18.2%**	13.1%	37.2%
	6	0.0%	0.0%	0.5%	0.5%	**4.7%**	0.9%	6.5%
	7	0.5%	0.2%	1.4%	0.9%	0.9%	1.4%	5.2%
	SUM	13.7%	11.9%	17.6%	7.0%	32.7%	17.1%	100.0%
POST		Tier-3 explanation post						
		1	**2**	**3**	**4**	**5**	**6**	SUM
Tier-2 direction post	**1**	6.4%	0.4%	3.1%	0.7%	2.6%	0.2%	13.5%
	2	0.9%	3.1%	1.3%	0.9%	1.1%	0.0%	7.3%
	3	0.4%	0.7%	2.4%	0.2%	2.2%	0.4%	6.4%
	4	0.4%	1.1%	2.2%	1.1%	4.6%	0.0%	9.5%
	5	0.4%	2.0%	2.2%	1.8%	**29.4%**	7.5%	43.3%
	6	1.3%	0.2%	1.3%	0.7%	**13.9%**	1.1%	18.5%
	7	0.0%	0.0%	0.0%	0.2%	0.9%	0.4%	1.5%
	SUM	9.9%	7.5%	12.6%	5.5%	54.7%	9.7%	100.0%

10.7.10.3 PR: A proton is moving to the right between two bar magnets

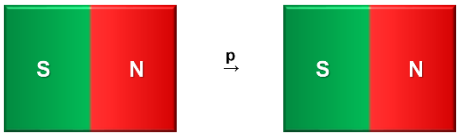
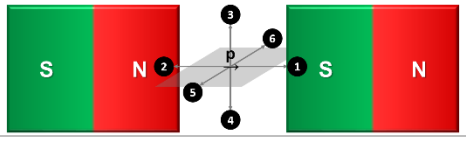
code	Text and picture	Answer options
pr_yn	**A proton is moving to the right at constant speed through the uniform magnetic field between two bar magnets.** **Do the magnets exert a force on the proton?**  1=Yes, the magnets exert a force on the proton. 2=No, the magnets do not exert a force on the proton.	
pr_dir	**If the magnets exert a force on the proton: In which direction does the force point?** The coordinate system below might help you to identify the direction. 	1=to the right (1). 2=to the left (2). 3=upwards (3). 4=downwards (4). 5=out of the page (5). 6=into the page (6). 7=in a different direction.
pr_expl	**Explanation:** 1=The south pole S of the magnets attracts the proton. 2=The north pole N of the magnets attracts the proton. 3=The magnetic force acts in the direction of the magnetic field. 4=The magnetic force acts in the direction of the movement of the proton. 5=The magnetic force acts on the proton according to the right-hand-rule. 6=None of the explanations above apply.	
pr_sure	**How confident are you that your answers are correct?** 1= not confident at all|very confident =5	

Table 138. Item PR: Students' answers in the pre- and post-visit questionnaire.

	N_{pre}	%	N_{post}	%
Tier-1: Do the magnets exert a force on the proton?	**485**		**509**	
Yes	259	53%	241	49%
No	226	47%	251	51%
Tier-2: If yes, which direction?	**246**		**257**	
1=to the right (1)	130	53%	88	34%
2=to the left (2)	62	25%	44	17%
3=upwards (3)	14	6%	42	16%
4=downwards (4)	19	8%	26	10%
5=out of the page (5)	7	3%	22	9%
6=into the page (6)	10	4%	25	10%
7=in a different direction	4	2%	10	4%

	N_{pre}	%	N_{post}	%
Tier-3: Explanation	**492**		**509**	
1=The south pole S of the magnets attracts the proton.	90	18%	55	11%
2=The north pole N of the magnets attracts the proton.	59	12%	43	8%
3=The magnetic force acts in the direction of the magnetic field.	76	15%	61	12%
4=The magnetic force acts in the direction of the movement of the proton.	82	17%	92	18%
5=The magnetic force acts on the proton according to the right-hand-rule.	64	13%	139	27%
6=None of the explanations above apply.	121	25%	119	23%
Tier-4: How confident are you that your answers are correct?	**492**		**509**	
1= not confident at all	110	22%	76	15%
2	141	29%	120	24%
3	116	24%	142	28%
4	76	15%	102	20%
5=very confident	49	10%	69	14%
Average confidence		**40%**		**49%**

Table 139. Item PR: Combinations of tier-2 and tier-3 level answer choices during the pre- and post-visit questionnaire. Combination frequencies larger than two times the guessing ratio are highlighted in orange. The correct combination is marked in blue, answers that were also rated correct are marked in green.

PRE		Tier-3 explanation pre						
		1	**2**	**3**	**4**	**5**	**6**	SUM
Tier-2 direction pre	**1**	25.6%	2.0%	15.4%	8.1%	1.2%	0.4%	52.8%
	2	1.6%	15.9%	2.4%	1.6%	2.4%	1.2%	25.2%
	3	0.8%	0.4%	1.2%	0.4%	2.0%	0.8%	5.7%
	4	1.2%	0.4%	0.8%	1.2%	3.7%	0.4%	7.7%
	5	0.0%	0.4%	0.8%	0.0%	1.6%	0.0%	2.8%
	6	0.8%	0.0%	1.2%	0.4%	1.2%	0.4%	4.1%
	7	0.0%	0.0%	0.8%	0.0%	0.0%	0.8%	1.6%
	SUM	30.1%	19.1%	22.8%	11.8%	12.2%	4.1%	100.0%
POST		Tier-3 explanation post						
		1	**2**	**3**	**4**	**5**	**6**	SUM
Tier-2 direction post	**1**	16.3%	0.8%	7.0%	4.3%	5.8%	0.0%	34.2%
	2	1.9%	10.1%	3.1%	0.8%	0.8%	0.4%	17.1%
	3	0.4%	1.6%	2.3%	1.2%	8.9%	1.9%	16.3%
	4	0.8%	1.6%	0.8%	2.7%	3.9%	0.4%	10.1%
	5	0.0%	0.4%	1.6%	1.9%	4.3%	0.4%	8.6%
	6	0.4%	0.4%	1.2%	1.2%	5.8%	0.8%	9.7%
	7	0.0%	0.4%	0.0%	0.8%	1.2%	1.6%	3.9%
	SUM	19.8%	15.2%	16.0%	12.8%	30.7%	5.4%	100.0%

10.7.10.4 PS: A proton is at rest between two magnets

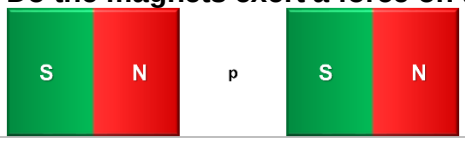
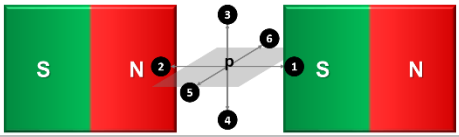
code	Text and picture	Answer options
ps_yn	**A proton is at rest in the uniform magnetic field between two bar magnets.** **Do the magnets exert a force on the proton?** 	1=Yes, the magnets exert a force on the proton. 2=No, the magnets do not exert a force on the proton.
ps_dir	**If the magnets exert a force on the proton: In which direction does the force point?** The coordinate system below might help you to identify the direction. 	1=to the right (1). 2=to the left (2). 3=upwards (3). 4=downwards (4). 5=out of the page (5). 6=into the page (6). 7=in a different direction.
ps_expl	**Explanation:** 1=The south pole S of the magnets attracts the proton. 2=The north pole N of the magnets attracts the proton. 3=The magnetic force acts in the direction of the magnetic field. 4=The magnetic force acts in the direction of the movement of the proton. 5=The magnetic force acts on the proton according to the right-hand-rule. 6=None of the explanations above apply.	
pr_sure	**How confident are you that your answers are correct?** 1= not confident at all|very confident =5	

Table 140. Item PS: Students' answers in the pre- and post-visit questionnaire.

	N_{pre}	%	N_{post}	%
Tier-1: Do the magnets exert a force on the proton?	**473**			
Yes	218	46%	187	38%
No	255	54%	305	62%
Tier-2: If yes, which direction?	**214**		**204**	
1=to the right (1)	99	46%	78	38%
2=to the left (2)	52	24%	44	22%
3=upwards (3)	16	7%	15	7%
4=downwards (4)	10	5%	18	9%
5=out of the page (5)	7	3%	24	12%
6=into the page (6)	7	3%	9	4%
7=in a different direction	23	11%	16	8%
Tier-3: Explanation	**492**		**509**	
1=The south pole S of the magnets attracts the proton.	69	14%	55	11%
2=The north pole N of the magnets attracts the proton.	53	11%	45	9%
3=The magnetic force acts in the direction of the magnetic field.	68	14%	49	10%
4=The magnetic force acts in the direction of the movement of the proton.	30	6%	46	9%
5=The magnetic force acts on the proton according to the right-hand-rule.	37	8%	90	18%
6=None of the explanations above apply.	235	48%	224	44%

	N_{pre}	%	N_{post}	%
Tier-4: How confident are you that your answers are correct?	**492**		**509**	
1= not confident at all	118	24%	100	20%
2	135	27%	117	23%
3	118	24%	137	27%
4	66	13%	90	18%
5=very confident	55	11%	65	13%
Average confidence		**40%**		**46%**

Table 141. Item PS: Combinations of tier-2 and tier-3 level answer choices during the pre- and post-visit questionnaire. Combination frequencies larger than two times the guessing ratio are highlighted in orange. The correct combination is marked in blue, answers that were also rated correct are marked in green.

PRE		Tier-3 explanation pre						
		1	**2**	**3**	**4**	**5**	**6**	SUM
Tier-2 direction pre	**1**	24.8%	0.9%	17.3%	1.9%	1.4%	0.0%	46.3%
	2	1.4%	18.7%	0.9%	1.4%	1.9%	0.0%	24.3%
	3	1.9%	0.9%	1.4%	0.9%	0.9%	1.4%	7.5%
	4	0.0%	0.0%	1.9%	0.5%	1.9%	0.5%	4.7%
	5	0.0%	0.5%	0.9%	0.0%	1.4%	0.5%	3.3%
	6	0.9%	0.0%	0.5%	0.9%	0.5%	0.5%	3.3%
	7	0.0%	0.0%	0.5%	0.9%	0.9%	8.4%	10.7%
	SUM	29.0%	21.0%	23.4%	6.5%	8.9%	11.2%	100.0%
POST		Tier-3 explanation post						
		1	**2**	**3**	**4**	**5**	**6**	SUM
Tier-2 direction post	**1**	23.0%	1.0%	7.4%	2.5%	3.4%	1.0%	38.2%
	2	1.0%	14.2%	3.4%	0.5%	2.0%	0.5%	21.6%
	3	0.5%	1.5%	2.5%	0.5%	2.5%	0.0%	7.4%
	4	1.5%	1.0%	0.5%	2.9%	2.0%	1.0%	8.8%
	5	0.5%	2.0%	3.4%	1.5%	3.9%	0.5%	11.8%
	6	0.0%	0.5%	0.0%	1.0%	2.0%	1.0%	4.4%
	7	0.0%	0.0%	0.0%	2.5%	1.5%	3.9%	7.8%
	SUM	26.5%	20.1%	17.2%	11.3%	17.2%	7.8%	100.0%

10.7.10.5 Grading of items related to the cloud chamber workshop CM, CS

Code	Formulae	pre		post	
		M	SD	M	SD
cm_1	1: AND(cm_yn=1;OR(cm_exp=4;cm_exp=5))	0.26	0.44	0.43	0.50
cs_1	1: AND(cs_yn=1;OR(cs_exp=4;cs_exp=5))	0.38	0.49	0.51	0.50

10.7.10.6 CM: Detector in the valley vs. on top of a mountain at 5000 m altitude

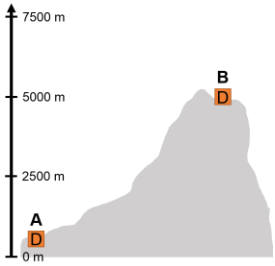
code	Text and picture	Answer options
cm_yn	**Imagine you compare two particle detectors [D]. One is placed in the valley at 0 m altitude. The other one is placed on top of a mountain at 5000 m altitude.** **Will the particle detector on top of the mountain detect a higher number of particles than the detector at 0 m?**  1=Yes, the particle detector on top of the mountain will detect a HIGHER number of particles. 2=No, the particle detector on top of the mountain will detect a LOWER number of particles. 3=No, there won't be a difference.	
cm_exp	**Explanation:** 1=The properties of the atmosphere on top of the mountain and in the valley are the same. 2=The properties of the ground on top of the mountain and in the valley are the same. 3=The Earth is the main source of the detected particles. 4=Space is the main source of the detected particles. 5=The atmosphere has a shielding effect. 6=There are fewer air molecules on top of the mountain. !other=different explanation:	
cm_sure	**How confident are you that your answers are correct?**	1= not confident at all 5= very confident

Table 142. Item CM: Students' answers in the pre- and post-visit questionnaire.

	N_{pre}	%	N_{post}	%
Tier-1: Will the particle detector on top of the mountain detect a higher number of particles than the detector at 0 m?	**509**		**509**	
1=Yes, the particle detector on top of the mountain will detect a HIGHER number of particles.	190	37%	273	54%
2=No, the particle detector on top of the mountain will detect a LOWER number of particles.	245	48%	149	29%
3=No, there won't be a difference.	74	15%	87	17%
Tier-2: Explanation	**452**		**509**	
1=The properties of the atmosphere on top of the mountain and in the valley are the same.	39	9%	51	10%
2=The properties of the ground on top of the mountain and in the valley are the same.	21	5%	24	5%
3=The Earth is the source of the detected particles.	32	7%	43	8%
4=Space is the source of the detected particles.	87	19%	189	37%
5=The atmosphere has a shielding effect.	70	15%	88	17%
6=There are fewer air molecules on top of the mountain.	183	40%	100	20%
7=different explanation	20	4%	14	3%
Tier-3: How confident are you that your answers are correct?	**492**		**509**	
1= not confident at all	86	17%	61	12%
2	120	24%	85	17%
3	154	31%	154	30%
4	93	19%	141	28%
5=very confident	39	8%	68	13%
Average confidence		**44%**		**54%**

Table 143. Item CM: Combinations of tier-1 and tier-2 level answer choices during the pre- and post-visit questionnaire. Combination frequencies larger than two times the guessing ratio are highlighted in orange. The correct combination is marked in blue, answers that were also rated correct are marked in green.

PRE		Tier-2 explanation pre							
		1	**2**	**3**	**4**	**5**	**6**	7	SUM
Tier-1 direction pre	**1**	1.1%	0.4%	1.8%	16.2%	11.1%	2.9%	2.2%	35.6%
	2	0.2%	0.9%	4.6%	1.1%	2.9%	37.4%	1.5%	48.7%
	3	7.3%	3.3%	0.7%	2.0%	1.5%	0.2%	0.7%	15.7%
	SUM	8.6%	4.6%	7.1%	19.2%	15.5%	40.5%	4.4%	100.0%
POST		Tier-2 explanation post							
		1	**2**	**3**	**4**	**5**	**6**		SUM
Tier-1 direction post	**1**	1.6%	1.4%	1.8%	31.6%	13.4%	2.4%	1.6%	53.6%
	2	1.0%	0.8%	5.3%	2.4%	2.9%	16.3%	0.6%	29.3%
	3	7.5%	2.6%	1.4%	3.1%	1.0%	1.0%	0.6%	17.1%
	SUM	10.0%	4.7%	8.4%	37.1%	17.3%	19.6%	2.8%	100.0%

10.7.10.7 CS: Detectors on the surface of the Earth vs. on a satellite

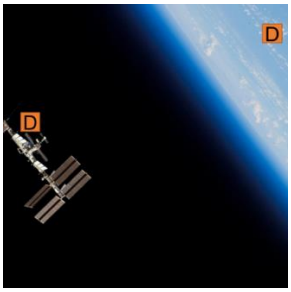
code	Text and picture	Answer options
cs_yn	**Imagine you compare two particle detectors [D]. One is placed on the surface of the Earth at 0 km altitude. The other one is placed on a satellite at 400 km altitude.** **Will the particle detector on the satellite detect a higher number of particles than the detector on the surface of the Earth?**  1=Yes, the particle detector on the satellite will detect a HIGHER number of particles. 2=No, the particle detector on the satellite will detect a LOWER number of particles. 3=No, there won't be a difference.	
cs_exp	**Explanation:** 1=The properties of the atmosphere at 400 km altitude and at 0 km altitude are the same. 3=The Earth is the main source of the detected particles. 4=Space is the main source of the detected particles. 5=The atmosphere has a shielding effect. 6=There are fewer air molecules in space. !other=different explanation:	
cs_sure	**How confident are you that your answers are correct?**	1= not confident at all 5= very confident

Table 144. Item CS: Students' answers in the pre- and post-visit questionnaire.

	N_{pre}	%	N_{post}	%
Tier-1: Will the particle detector on the satellite detect a higher number of particles than the detector on the surface of the Earth?	**508**		**492**	
1=Yes, the particle detector on the satellite will detect a HIGHER number of particles.	238	47%	314	64%
2=No, the particle detector on the satellite will detect a LOWER number of particles.	252	50%	153	31%
3=No, there won't be a difference.	18	4%	25	5%
Tier-2: Explanation	**450**		**505**	
1=The properties of the atmosphere at 400 km altitude and 0 km altitude are the same.	13	3%	20	4%
3=The Earth is the source of the detected particles.	73	16%	76	15%
4=Space is the source of the detected particles.	110	24%	195	38%
5=The atmosphere has a shielding effect.	105	23%	130	26%
6=There are fewer air molecules in space.	125	28%	69	14%
7=different explanation	24	5%	15	3%
Tier-3: How confident are you that your answers are correct?	**492**		**509**	
1= not confident at all	88	18%	73	14%
2	127	26%	81	16%
3	139	28%	140	28%
4	85	17%	129	25%
5=very confident	53	11%	86	17%
Average confidence		**44%**		**54%**

Table 145. Item CS: Combinations of tier-1 and tier-2 level answer choices during the pre- and post-visit questionnaire. Combination frequencies larger than two times the guessing ratio are highlighted in orange. The correct combination is marked in blue, answers that were also rated correct are marked in green.

PRE		Tier-2 explanation pre							
		1	**2**	**3**	**4**	**5**	**6**	7	SUM
Tier-1 direction pre	**1**	0.2%	0.0%	2.2%	21.7%	17.9%	0.7%	2.4%	45.1%
	2	1.1%	0.4%	13.3%	1.8%	5.1%	26.8%	2.9%	51.3%
	3	1.5%	0.0%	0.7%	0.9%	0.2%	0.2%	0.0%	3.5%
	SUM	2.9%	0.4%	16.2%	24.3%	23.2%	27.7%	5.3%	100.0%
POST		Tier-2 explanation post							
		1	**2**	**3**	**4**	**5**	**6**		SUM
Tier-1 direction post	**1**	0.6%	0.0%	5.7%	34.8%	20.9%	1.2%	0.6%	63.8%
	2	0.6%	0.8%	7.5%	3.0%	5.1%	12.4%	1.6%	31.1%
	3	2.4%	0.0%	0.8%	0.4%	0.2%	0.4%	0.8%	5.1%
	SUM	3.7%	0.8%	14.0%	38.2%	26.2%	14.0%	3.0%	100.0%

10.7.10.8 XC: Dangers for the cleaning staff of a hospital

code	Text and picture	Answer options
xc	**Is it dangerous for the cleaning staff of a hospital when they clean a room with an X-ray machine?** 1=Yes, parts of the X-ray device are exposed to X-rays and therefore become radioactive themselves. 2=Yes, because X-ray machines are always a source of radiation, even if they are switched off. 3=No, there is no danger, because the X-ray units are switched off during the cleaning. 4=No, provided that the entire air of the treatment rooms has been exchanged before the cleaning. 5=No, provided that the cleaning staff waits a while before they enter the room.	
xc_sure	**How confident are you that your answers are correct?**	1= not confident at all 5= very confident

Table 146. Item XC: Students' answers in the pre- and post-visit questionnaire.

	N_{pre}	%	N_{post}	%
Tier-1: Is it dangerous for the cleaning staff of a hospital when they clean a room with an X-ray machine?	**509**		**509**	
1=Yes, parts of the device are exposed to X-rays and therefore become radioactive themselves.	56	11%	27	5%
2=Yes, because of X-ray machines are always a source of radiation, even if they are switched off.	97	19%	30	6%
3=No, there is no danger, because the X-ray units are switched off during the cleaning.	266	52%	375	74%
4=No, provided that the entire air of the treatment rooms has been exchanged before the cleaning.	37	7%	24	5%
5=No, provided that the cleaning staff waits a while before they enter the room.	53	10%	53	10%
Tier-2: How confident are you that your answers are correct?	**492**		**509**	
1= not confident at all	68	14%	27	5%
2	103	21%	47	9%
3	176	36%	106	21%
4	98	20%	164	32%
5=very confident	47	10%	165	32%
Average confidence		**48%**		**69%**

10.7.10.9 XW: Danger of wearing a watch that was X-rayed for 1 hour

code	Text and picture	Answer options
xw	**A watch is X-rayed for 1 hour inside an X-ray machine. Is it dangerous to wear the watch afterwards?** 1=Yes, ionizing radiation accumulates in the watch, therefore the watch will emit radiation afterwards. 2=Yes, the watch has been exposed to X-rays and therefore the atoms in the watch become radioactive. 3=No, to become radioactive itself, the watch needs to be exposed to X-rays for a longer time. 4=No, as long as no radioactive atoms are transferred to the watch, the watch does not emit radiation.	
xw_sure	**How confident are you that your answers are correct?**	1= not confident at all 5= very confident

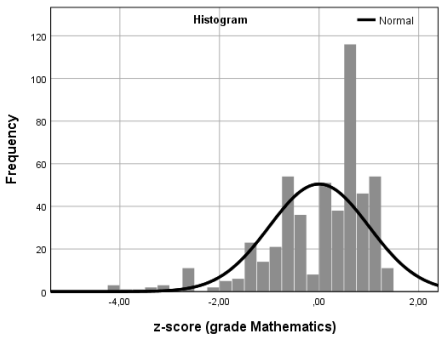
Table 147. Item XW: Students' answers in the pre- and post-visit questionnaire.

	N_{pre}	%	N_{post}	%
Tier-1: A watch is X-rayed for 1 hour inside an X-ray machine. Is it dangerous to wear the watch afterwards?	**509**		**509**	
1=Yes, ionizing radiation accumulates in the watch, therefore the watch will emit radiation afterwards.	175	34%	51	10%
2=Yes, the watch has been exposed to X-rays and therefore the atoms in the watch become radioactive.	79	16%	32	6%
3=No, to become radioactive itself, the watch needs to be exposed to X-rays for a longer time.	67	13%	92	18%
4=No, as long as no radioactive atoms are transferred to the watch, the watch does not emit radiation.	188	37%	334	66%
Tier-2: How confident are you that your answers are correct?	**492**		**509**	
1= not confident at all	67	14%	31	6%
2	155	32%	64	13%
3	141	29%	138	27%
4	100	20%	153	30%
5=very confident	29	6%	123	24%
Average confidence		**43%**		**64%**

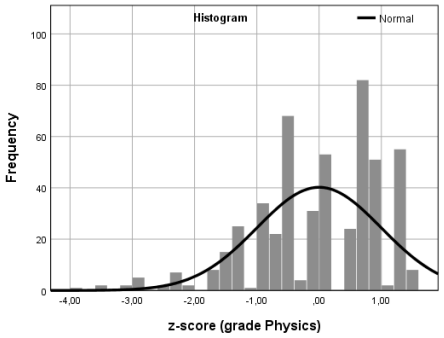
10.7.11 School grades (MA, PH, EN grade)

Extreme cases	Extreme values	Handling
101	z(MA)=-4.2	Due to the deviation from a normal distribution, extreme values were not handled as outliers. But Mathematics grade was discarded as predictor. English grades were dichotomised, see 7.3.10.
133	z(MA)=-3.6, z(PH)=-3.4	
191	z(MA)=-3.9	
196	z(MA)=-4.2, z(PH)=-3.4	
209	z(ENgrade)=-3.9	
282	z(ENgrade)=-3.6	
354	z(MA)=-4.2, z(PH)=-4.0	

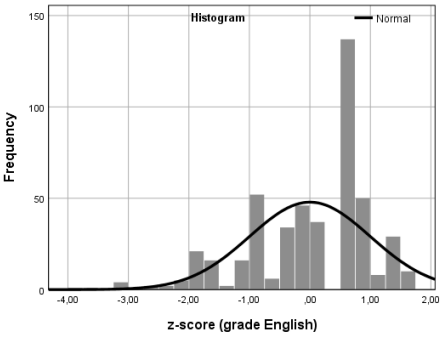
z-score: Mathematics grade (Z_MA)	
Skewness	**-1.37**
SE	0.11
z-value	12>>3.29, significant
Kurtosis	**2.43**
SE	0.22
z-value	11>>3.29, significant
Kolmo	KS(506)= 0.16, p < 0.001
Shapiro	SW(506) = 0.89, p < 0.001



z-score: Physics grade (Z_PH)	
Skewness	**-0.86**
SE	0.11
z-value	8>>3.29, significant
Kurtosis	**0.67**
SE	0.22
z-value	3<3.29, n.s.
Kolmo	KS(504)= 0.14, p < 0.001
Shapiro	SW(504) = 0.93, p < 0.001



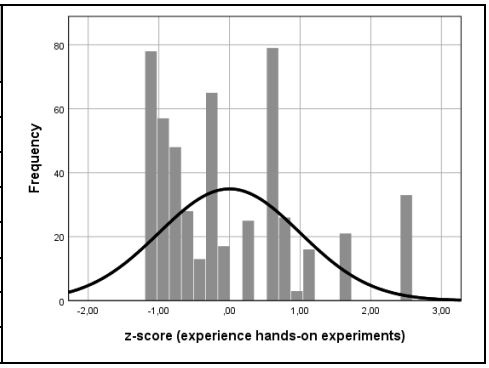
z-score: English grade (Z_ENgrade)	
Skewness	**-0.95**
SE	0.11
z-value	9 >> 3.29, significant
Kurtosis	**0.77**
SE	0.22
z-value	3.5 > 3.29, significant
Kolmo	KS(480)= 0.19, p < 0.001
Shapiro	SW(480) = 0.92, p < 0.001



10.7.12 Experience with hands-on experiments (EH)

No extreme cases

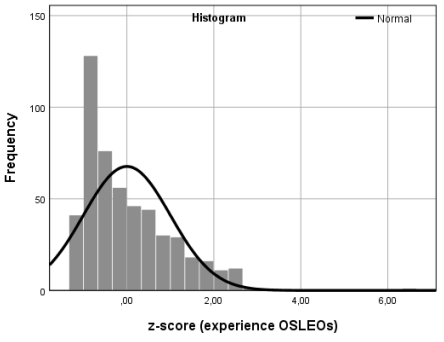
z-score: experience hands-on experiments (Z_EH)	
Skewness	**0.92**
SE	0.11
z-value	8 > 3.29, significant
Kurtosis	**0.11**
SE	0.22
z-value	0.5 < 3.29, n.s.
Kolmo	KS(509)= 0.15, p < 0.001
Shapiro	SW(509) = 0.88, p < 0.001



10.7.13 Experience with out-of-school learning opportunities (EO)

Extreme case	Extreme value	Handling
72	z(EO)= 5.1	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
319	z(EO)= 3.8	
365	z(EO)= 4.2	
367	z(EO)= 3.8	
385	z(EO)= 4.1	
389	z(EO)= 3.5	
439	z(EO)= 3.9	
456	z(EO)= 3.5	
410	z(EO)= 13.4	Handled as outlier, set to 5.1 as nearest value

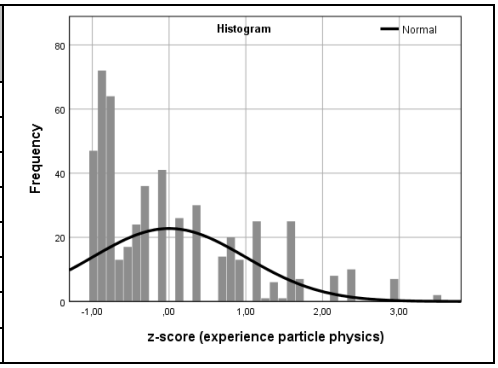
z-score: experience out-of-school learning opportunities (Z_EO)	
Skewness	**1.33**
SE	0.11
z-value	12 >> 3.29, significant
Kurtosis	**3.10**
SE	0.22
z-value	14 >> 3.29, significant
Kolmo	KS(509)= 0.13, p < 0.001
Shapiro	SW(509) = 0.88, p < 0.001



10.7.14 Experience with particle physics (EP)

Extreme cases	Extreme values	Handling
257	z(EP)=3.5	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
448	z(EP)=3.5	

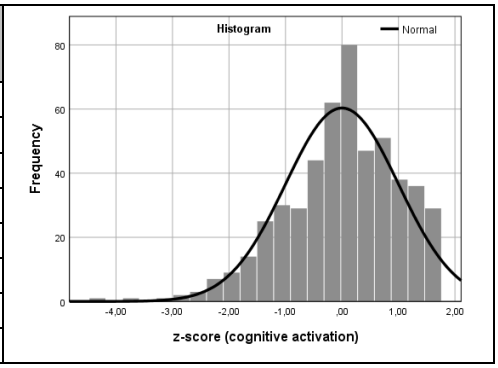
z-score: experience with particle physics (Z_EP)	
Skewness	**1.08**
SE	0.11
z-value	10 > 3.29, significant
Kurtosis	**0.44**
SE	0.22
z-value	2 < 3.29, n.s.
Kolmo	KS(509)= 0.18, p < 0.001
Shapiro	SW(509) = 0.86, p < 0.001



10.7.15 Cognitive activation (CA)

Extreme cases	Extreme values	Handling
130	z(CA)=-4.3	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
7	Z(CA)=-3.7	

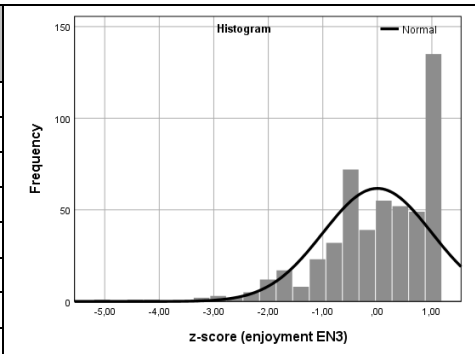
z-score: cognitive activation (Z_CA)	
Skewness	**-0.63**
SE	0.11
z-value	6 > 3.29, significant
Kurtosis	**0.59**
SE	0.22
z-value	2.7 < 3.29, n. s.
Kolmo	KS(509)= 0.11, p < 0.001
Shapiro	SW(509) = 0.96, p < 0.001



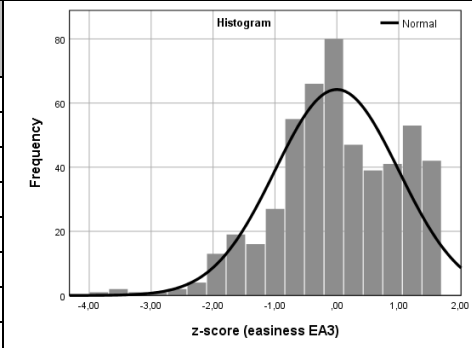
10.7.16 Enjoyment (ENJ), Easiness (EAS) & Helpfulness (HEL)

Extreme cases	Extreme values	Handling
328	z(HEL)=-3.7	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
370	z(EAS)=-3.5	
146	z(ENJ)=-5.0, z(EAS)=-3.8	
266	z(ENJ)=-4.1	
363	z(ENJ)=-4.4, z(EAS)=-3.5	
167	z(HEL)=-3.5	

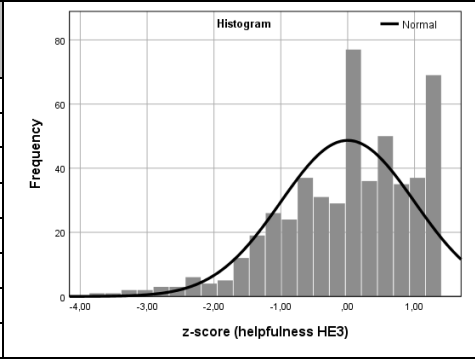
z-score: enjoyment (Z_ENJ)	
Skewness	**-1.21**
SE	0.11
z-value	11 >> 3.29, significant
Kurtosis	**2.14**
SE	0.22
z-value	10 > 3.29, significant
Kolmo	KS(509)= 0.15, p < 0.001
Shapiro	SW(509) = 0.88, p < 0.001



z-score: easiness (Z_EAS)	
Skewness	**-0.51**
SE	0.11
z-value	5 > 3.29, significant
Kurtosis	**0.34**
SE	0.22
z-value	1.5 < 3.29, n. s.
Kolmo	KS(509)= 0.08, p < 0.001
Shapiro	SW(509) = 0.96, p < 0.001



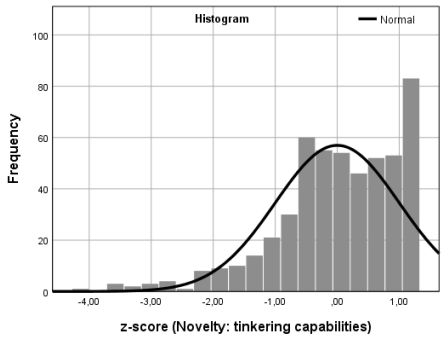
z-score: helpfulness (Z_HEL)	
Skewness	**-0.83**
SE	0.11
z-value	8 > 3.29, significant
Kurtosis	**0.59**
SE	0.22
z-value	2.6 < 3.29, n. s.
Kolmo	KS(509)= 0.13, p < 0.001
Shapiro	SW(509) = 0.94, p < 0.001



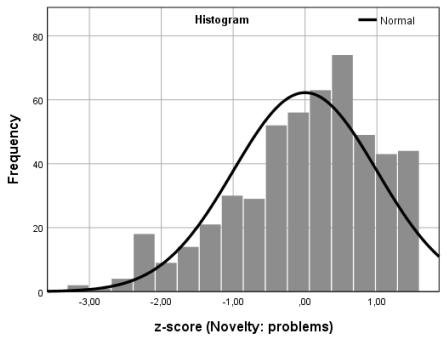
10.7.17 Tinkering support, lack of problems & setting orientation

Extreme cases	Extreme values	Handling
305	z(TIN)=-4.1	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
157	z(TIN)=-3.6	
153	z(TIN)=-3.6	
57	z(TIN)=-3.6	
363	z(TIN)=-3.3	

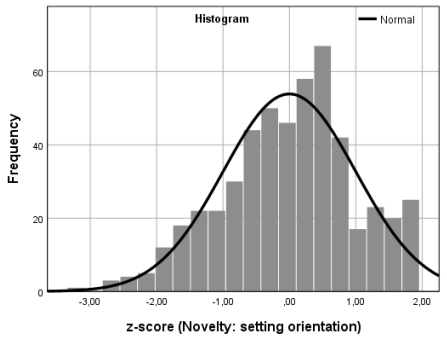
z-score: tinkering support (Z _TIN)	
Skewness	**-1.05**
SE	0.11
z-value	10 >> 3.29, significant
Kurtosis	**1.34**
SE	0.22
z-value	6 > 3.29, significant
Kolmo	KS(509)= 0.12, p < 0.001
Shapiro	SW(509) = 0.91, p < 0.001



z-score: lack of problems (Z_ PRO)	
Skewness	**-0.67**
SE	0.11
z-value	6 > 3.29, significant
Kurtosis	**-0.03**
SE	0.22
z-value	0 < 3.29, n. s.
Kolmo	KS(509)= 0.12, p < 0.001
Shapiro	SW(509) = 0.95, p < 0.001



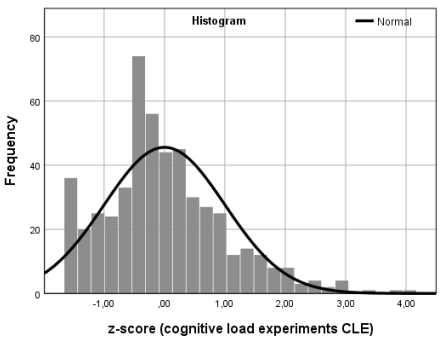
z-score: setting orientation (Z_SO)	
Skewness	**-0.28**
SE	0.11
z-value	2.5 < 3.29, n. s.
Kurtosis	**-0.21**
SE	0.22
z-value	1.0 < 3.29, n. s.
Kolmo	KS(509)= 0.09, p < 0.001
Shapiro	SW(509) = 0.98, p < 0.001



10.7.18 Cognitive load experiments (CLE)

Extreme cases	Extreme values	Handling
133	z(CLE)=4.1	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
253	Z(CLE)=3.8	
378	Z(CLE)=3.4	

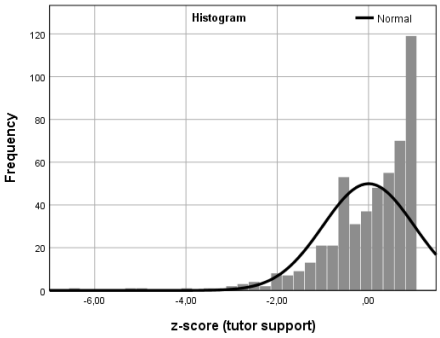
z-score: cognitive load experiments (Z_CLE)	
Skewness	**0.80**
SE	0.11
z-value	8 > 3.29, significant
Kurtosis	**0.94**
SE	0.22
z-value	4 > 3.29, significant
Kolmo	KS(509)= 0.11, p < 0.001
Shapiro	SW(509) = 0.95, p < 0.001



10.7.19 Tutor support (TS)

Extreme cases	Extreme values	Handling
187	z(TS)=-6.4	Due to the deviation from a normal distribution, extreme values were not handled as outliers.
320	z(TS)=-5.2	
146	z(TS)=-5.0	
294	z(TS)=-4.0	
85	z(TS)=-3.5	

z-score: tutor support (Z_TS)	
Skewness	**-1.89**
SE	0.11
z-value	17 >> 3.29, significant
Kurtosis	**6.06**
SE	0.22
z-value	28 >> 3.29, significant
Kolmo	KS(509)= 0.11, p < 0.001
Shapiro	SW(509) = 0.96, p < 0.001



10.8 Summary of extreme values and multivariate outliers

case	student	teacher	group	country	f/m	Univariate extreme values	Multivariate outlier[51]
7	T11F111	67	1	CH	m	z(CA)=-3.7	
29	K011a212	60	3	PL	m		SC3
57	N11A111	59	4	ES	m	z(NOV_TIN)=-3.6	
72	J010J1	56	5	CZ	m	z(EO)= 5.1	IE3, SC3, CS3, CA, UN3, ENJ, HEL, EAS
85	B03O0	52	6	UK	m	z(TS)=-3.5	CS3, CA, UN3, ENJ, HEL, EAS
101	R011S0	52	6	UK	f	z(MA)=-4.2	SC3
130	A03C0	46	8	AT	m	z(CA)=-4.3	CA*
133	B210S0	46	8	AT	f	z(MA)=-3.6, z(PH)=-3.4, z(IE1)=-3.4, z(CLE)=4.1	IE3, CS3, ENJ, HEL
137	G06C0	46	8	AT	m	z(SC1)=-3.3, z(SCE1)=-3.5	
146	M01M1	46	8	AT	m	z(IE1)=-3.4, z(IE3)=-3.8, z(SC1)=-3.3, z(SCE1)=-3.8, z(SCE2)=-3.6, z(CS2)=-3.4, z(CS3)=-3.4, z(EN3)=-5.0, z(EA3)=-3.8, z(TS)=-5.0	IE3, IE3*, SC3, SC3*, CS3, ENJ, HEL, EAS
153	S012D0	46	8	AT	f	z(NOV_TIN)=-3.6	
157	E011S012	45	9	DE	m	z(CS1)=-4.1, z(NOV_TIN)=-3.6	IE3, SC3, CS3, CA, UN3, EAS
161	L03B012	45	9	DE	f		EN*
167	R011L0	45	9	DE	m	z(HE3)=-3.5	HE*
187	S05L0	44	10	NO	f	z(SC3)=-4.1, z(SCE3)=-4.8, z(TS)=-6.4	IE3, SC3, SC*, CS3, ENJ, HEL, EAS
191	T011B1	44	10	NO	m	z(MA)=-3.9, z(SCE1)=-3.5	
196	D09M0	40	11	ES	m	z(MA)=-4.2, z(PH)=-3.4, z(IE1)=-3.4, z(SC1)=-3.3, z(CS1)=-3.4	
198	D13C013	40	11	ES	m		CS3*
201	H010A0	40	11	ES	m		CS3
209	H07Q0	38	12	PT	f	z(EN)=-3.9	
253	E11D0	31	14	IT	m	z(CLE)=3.8	IE3, SC3, CS3, CA, UN3, ENJ, HEL, EAS
257	F06L0	28	14	IT	m	z(EP)=3.5	
266	G15L0	30	14	IT	m	z(EN3)=-4.1	EN*
269	L010A0	27	14	IT	m	z(CT1)=-6.2, z(CS1)=-4.1	IE3, CS3

[51] Multivariate outliers were identified using the Mahalanobis distance. Values marked with an * were detected because they showed extreme residuals.

case	student	teacher	group	country	f/m	Univariate extreme values	Multivariate outlier
282	B02A1	26	15	FR	m	z(EN)=-3.6	
287	J01F1	26	15	FR	m		CS3*
294	M22V1	26	15	FR	f	z(SCE3)=-3.5, z(TS)=-4.0	
305	H07A0	21	16	CH	m	z(CS2)=-3.4, z(NOV_TIN)=-4.1	IE3*
319	A08M3	20	16	CH	m	z(EO)= 3.8	
320	A13I0	20	17	CH	m	z(SCE3)=-3.5, z(TS)=-5.2	IE3, SC3. CS3, CA, UN3, ENJ, HEL, EAS
323	K111D0	20	17	CH	m	z(CS2)=-3.4	
326	P03S1	20	17	CH	f	z(IE1)=-3.4, z(IE3)=-4.4, z(SC1)=-3.3, z(SCE1)=-3.8, z(CS1)=-4.1, z(CS2)=-3.4, z(CS3)=-3.9	
327	R012E0	20	17	CH	m	z(IE1)=-3.4, z(CS2)=-4.2	
328	S111D1	20	17	CH	m	z(HE3)=-3.7	IE3, SC3, CS3, ENJ, HEL, HEL*, EAS
353	A011K0	17	19	UK	m	z(CT1)=-4.7, z(CS1)=-4.1	CS3
354	A16L0	17	19	UK	m	z(MA)=-4.2, z(PH)=-4.0	UN3
363	M110F0	17	19	UK	m	z(EN3)=-4.4, z(EA3)=-3.5, z(NOV_TIN)=-3.3	EN*
365	O04D0	17	19	UK	m	z(EO)= 4.2	IE3, SC3, CS3, ENJ, HEL, EAS
367	T08C0	17	19	UK	m	z(EO)= 3.8	
370	B111N0	16	20	ES	m	z(EA3)=-3.5	
378	I012R1	16	20	ES	m	z(CLE)=3.4	IE3, SC3*, CS3, CA, UN3, ENJ, HEL, EAS
380	J05A1	16	20	ES	m	z(CT1)=-4.7	CS3
383	M03M012	16	20	ES	f	z(IE3)=-3.8	IE3*, CA*
385	M08E0	16	20	ES	m	z(EO)= 4.1	UN3, CA
386	M110A0	16	20	ES	m	z(CS3)=-3.9	CS3*
389	O011R0	16	20	ES	m	z(EO)= 3.5	
394	C08H0	15	21	SG	m		CA, UN3, ENJ, HEL, EAS
398	J06R0	15	21	SG	m	z(CT1)=-4.7	CS3
410	Y03W0	13	21	SG	m	z(EO)= 13.4 >> set to 5.1	IE3, SC3, CS3, CA, UN3, ENJ, HEL, EAS
434	E03S0	11	23	FR	f	z(SC3)=-4.8, z(SCE3)=-4.8	IE3, SC3*, CS3*
439	L06N1	11	23	FR	f	z(EO)= 3.9	IE3, SC3, CS3, ENJ, HEL, EAS
448	T17N0	11	23	FR	m	z(EP)=3.5	HEL*
456	G112A0	9	24	UK	f	z(EO)= 3.5	EAS
478	S07M113	7	25	NO	f		CS3*

10.9 Sequential linear modelling: supporting material

10.9.2 SLM Situational interest

Correlation matrix

Table 148. Pearson correlation r between predictors DV situational interest IE3, significant correlations highlighted (Bonferroni correction m = 78, p < 0.0006).

		SEX	AGE	IE1	CS1	EH	EN	EO	EP	TIN	SO	PRO	CLE	TS	IE3
Gender SEX	r	1	-0.12	-0.14	-0.06	-0.01	0.02	-0.03	-0.03	0.07	0.02	-0.04	0.07	0.05	0.03
	p		0.008	0.001	0.159	0.810	0.734	0.523	0.541	0.110	0.685	0.359	0.122	0.294	0.525
Age AGE	r		1	0.08	0.01	**-0.16**	0.01	-0.01	0.13	0.05	0.14	0.10	-0.06	0.01	0.01
	p			0.075	0.794	2E-04	0.813	0.830	0.002	0.238	0.001	0.019	0.176	0.824	0.881
Dispositional Physics interest IE1	r			1	**0.65**	0.02	0.14	**0.16**	0.15	**0.31**	**0.40**	0.11	**-0.28**	**0.22**	**0.49**
	p				1E-63	0.724	0.002	2E-04	7E-04	7E-13	5E-21	0.018	2E-10	7E-07	3E-32
Curiosity state particle physics CS1	r				1	0.06	0.11	0.10	**0.17**	**0.38**	**0.34**	0.09	**-0.20**	**0.26**	**0.50**
	p					0.197	0.017	0.022	2E-04	3E-19	4E-15	0.050	4E-06	3E-09	4E-33
Experience hands-on experiments EH	r					1	0.01	0.13	-0.01	0.04	0.09	0.03	-0.03	0.04	-0.02
	p						0.843	0.003	0.817	0.334	0.034	0.434	0.440	0.415	0.666
English skills (binary) EN	r						1	-0.01	0.07	0.11	0.07	0.11	-0.10	0.03	0.10
	p							0.859	0.120	0.010	0.104	0.014	0.022	0.434	0.025
Experience with OSLEOs EO	r							1	0.10	0.01	0.12	-0.01	-0.06	0.09	0.08
	p								0.018	0.791	0.009	0.868	0.200	0.038	0.085
Experience with particle physics EP	r								1	0.10	**0.20**	0.08	-0.13	0.08	0.09
	p									0.022	4E-06	0.058	0.004	0.076	0.046
Tinkering support (novelty) TIN	r									1	**0.48**	**0.35**	**-0.42**	**0.57**	**0.59**
	p										2E-30	9E-16	4E-23	3E-44	3E-48
Setting orientation (novelty) SO	r										1	**0.16**	**-0.32**	**0.28**	**0.44**
	p											4E-04	3E-13	8E-11	6E-26
Lack of problems (novelty) PRO	r											1	**-0.59**	**0.26**	**0.17**
	p												1E-49	2E-09	2E-04
Cognitive load experiments CLE	r												1	**-0.32**	**-0.32**
	p													2E-13	2E-13
Tutor support TS	r													1	**0.53**
	p														2E-38

Linear model based only on novelty experience factors

A separate linear model was calculated with only novelty experience factors to estimate, whether novelty influence factors are indeed mediated by novelty experience factors (see Table 149). In total, this model accounts for almost half (45%) of the variance in the dependent variable situational interest, which is significantly more than what novelty influence factors alone account for (31%). The regression coefficients of the significant predictors tinkering and tutor support are slightly altered ($\boldsymbol{\beta_{w/o} = 0.37^{***} | 0.25^{***}}$), and an additional predictor, the novelty experience factor 'setting orientation' becomes significant ($\boldsymbol{\beta_{w/o} = 0.17^{***}}$). As the identified novelty experience factors share variance, the interpretation of their individual contributions is difficult. In particular, the novelty factors 'tinkering support' and 'tutor support' are highly correlated ($r = -0.57$). Apparently, the relationship between novelty influence factors and the educational outcome situational interest is mediated to a high degree by novelty experience factors, but novelty experience factors even model additional variance after controlling for novelty influence factors.

Table 149. Sequential regression of novelty experience factors on students' situational interest: Standardised regression coefficient and p-values for sample without outliers ($\boldsymbol{\beta_{w/o}, p_{w/o}}$), 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[52] are highlighted.

	w/o outliers			95% C. I.		
	$N = 493$			Lower	Upper	Adj.
Predictor	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)			0.04	0.01	0.51	
Tinkering support (novelty) TIN	**0.37**	***	1E-13	0.27	0.46	**0.37**
Setting orientation (novelty) SO	**0.17**	***	1E-05	0.09	0.24	**0.40**
Lack of problems (novelty) PRO	-0.11		0.01	-0.19	-0.03	0.40
Cognitive load experiments CLE	-0.10		0.03	-0.21	-0.01	0.41
Tutor support TS	**0.25**	***	2E-08	0.18	0.36	**0.45**
	$Adjusted\ R^2 = 0.45, SEE = 0.72$ $F(5{,}487) = 79.8, p = 5 \cdot 10^{-61}$					

[52] The Bonferroni correction reduces the required p-values with $m = 13$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

10.9.3 SLM Curiosity state post

Correlation matrix

Table 150. Pearson correlation r between predictors and DV curiosity state post CS3, significant correlations highlighted (Bonferroni correction m = 66, p < 0.0008).

		CT	CS1	IE1	EH	EN	EO	EP	TIN	SO	PRO	CLE	TS	CS3
Curiosity trait CT	r	1	**0.52**	**0.43**	0.02	0.09	0.13	0.09	**0.31**	**0.24**	0.14	**-0.20**	**0.24**	**0.38**
	p		1E-36	6E-24	0.583	0.033	0.003	0.036	1E-12	6E-08	0.002	4E-06	5E-08	6E-19
Curiosity state particle physics CS1	r		1	**0.65**	0.06	0.11	0.10	**0.17**	**0.38**	**0.34**	0.09	**-0.20**	**0.26**	**0.57**
	p			1E-63	0.197	0.017	0.022	2E-04	3E-19	4E-15	0.050	4E-06	3E-09	6E-46
Dispositional Physics interest IE1	r			1	0.02	0.14	**0.16**	**0.15**	**0.31**	**0.40**	0.11	**-0.28**	**0.22**	**0.48**
	p				0.724	0.002	2E-04	7E-04	7E-13	5E-21	0.018	2E-10	7E-07	3E-31
Experience hands-on experiments EH	r				1	0.01	0.13	-0.01	0.04	0.09	0.03	-0.03	0.04	0.03
	p					0.843	0.003	0.817	0.334	0.034	0.434	0.440	0.415	0.493
English skills (binary) EN	r					1	-0.01	0.07	0.11	0.07	0.11	-0.10	0.03	0.13
	p						0.859	0.120	0.010	0.104	0.014	0.022	0.434	0.004
Experience with OSLEOs EO	r						1	0.10	0.01	0.12	-0.01	-0.06	0.09	0.09
	p							0.018	0.791	0.009	0.868	0.200	0.038	0.041
Experience with particle physics EP	r							1	0.10	**0.20**	0.08	-0.13	0.08	0.11
	p								0.022	4E-06	0.058	0.004	0.076	0.012
Tinkering support (novelty) TIN	r								1	**0.48**	**0.35**	**-0.42**	**0.57**	**0.49**
	p									2E-30	9E-16	4E-23	3E-44	4E-32
Setting orientation (novelty) SO	r									1	**0.16**	**-0.32**	**0.28**	**0.37**
	p										4E-04	3E-13	8E-11	5E-18
Lack of problems (novelty) PRO	r										1	**-0.59**	**0.26**	**0.11**
	p											1E-49	2E-09	0.010
Cognitive load experiments CLE	r											1	**-0.32**	**-0.24**
	p												2E-13	2E-08
Tutor support TS	r												1	**0.40**
	p													3E-21

Scatterplot

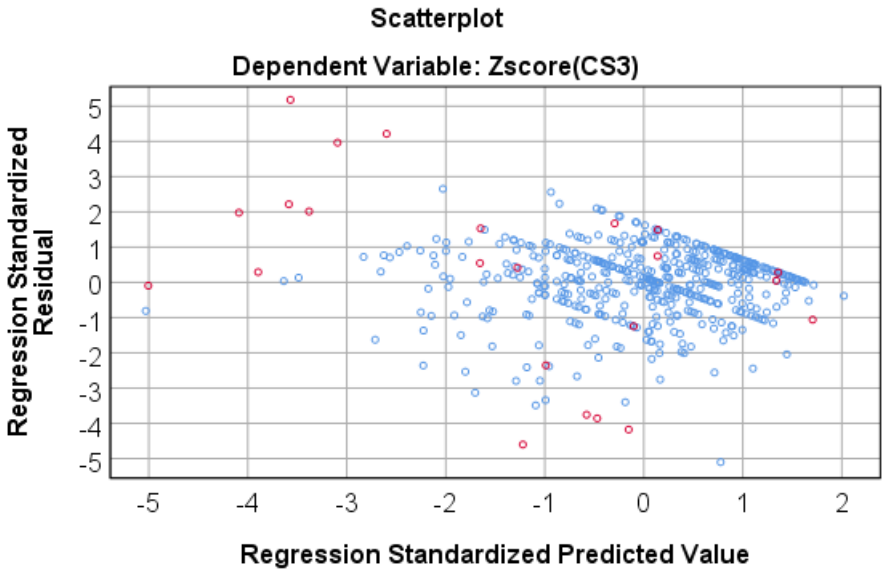


Figure 41. Residuals scatterplot for dependent variable curiosity state post (CS3). 23 cases displayed in red have been indentified as multivariate outliers (18 cases) or outliers in the solution (5 cases) and have been excluded from the analysis.

Linear model based only on novelty experience factors

Table 151. Sequential regression of novelty experience factors on students' curiosity state post: Standardised regression coefficient and p-values for sample without outliers $(\beta_{w/o}, p_{w/o})$, 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[53] are highlighted.

	w/o outliers			95% C. I.		
	$N = 495$			Lower	Upper	Adj.
Predictor	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)			0.618	-0.05	0.09	
Tinkering support (novelty) TIN	**0.29**	***	3E-07	0.18	0.39	**0.24**
Setting orientation (novelty) SO	**0.14**	*	0.002	0.05	0.22	**0.26**
Lack of problems (novelty) PRO	-0.10		0.031	-0.19	-0.01	0.26
Cognitive load experiments CLE	-0.07		0.190	-0.17	0.03	0.26
Tutor support TS	**0.21**	***	2E-05	0.12	0.34	**0.29**
	$Adjusted\ R^2 = 0.29, SEE = 0.80$ $F(5{,}480) = 40, p = 2 \cdot 10^{-34}$					

[53] The Bonferroni correction reduces the required p-values with $m = 12$ to * $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

10.9.4 SLMs Enjoyment and helpfulness: correlations matrix

Table 152. Pearson correlation r between predictors and DVs enjoyment ENJ & helpfulness HEL, sign. correlations highlighted (Bonferroni correction m = 65, p < 0.0008).

		IE1	EH	EN	EO	EP	TIN	SO	PRO	CLE	TS	ENJ	HEL
Dispositional Physics interest IE1	r	1	0.02	0.14	**0.16**	**0.15**	**0.31**	**0.40**	0.11	**-0.28**	**0.22**	**0.34**	**0.33**
	p		0.724	0.002	2E-04	7E-04	7E-13	5E-21	0.018	2E-10	7E-07	1E-15	2E-14
Experience hands-on experiments EH	r		1	0.01	0.13	-0.01	0.04	0.09	0.03	-0.03	0.04	0.01	0.09
	p			0.843	0.003	0.817	0.334	0.034	0.434	0.440	0.415	0.835	0.050
English skills (binary) EN	r			1	-0.01	0.07	0.11	0.07	0.11	-0.10	0.03	0.03	0.02
	p				0.859	0.120	0.010	0.104	0.014	0.022	0.434	0.457	0.692
Experience with OSLEOs EO	r				1	0.10	0.01	0.12	-0.01	-0.06	0.09	0.05	0.06
	p					0.018	0.791	0.009	0.868	0.200	0.038	0.240	0.182
Experience with particle physics EP	r					1	0.10	**0.20**	0.08	-0.13	0.08	0.07	-0.02
	p						0.022	4E-06	0.058	0.004	0.076	0.104	0.699
Tinkering support (novelty) TIN	r						1	**0.48**	**0.35**	**-0.42**	**0.57**	**0.64**	**0.52**
	p							2E-30	9E-16	4E-23	3E-44	7E-59	3E-37
Setting orientation (novelty) SO	r							1	**0.16**	**-0.32**	**0.28**	**0.38**	**0.38**
	p								4E-04	3E-13	8E-11	9E-19	3E-19
Lack of problems (novelty) PRO	r								1	**-0.59**	**0.26**	**0.24**	0.04
	p									1E-49	2E-09	4E-08	0.315
Cognitive load experiments CLE	r									1	**-0.32**	**-0.35**	**-0.18**
	p										2E-13	6E-16	5E-05
Tutor support TS	r										1	**0.55**	**0.42**
	p											3E-42	3E-23

Scatterplots

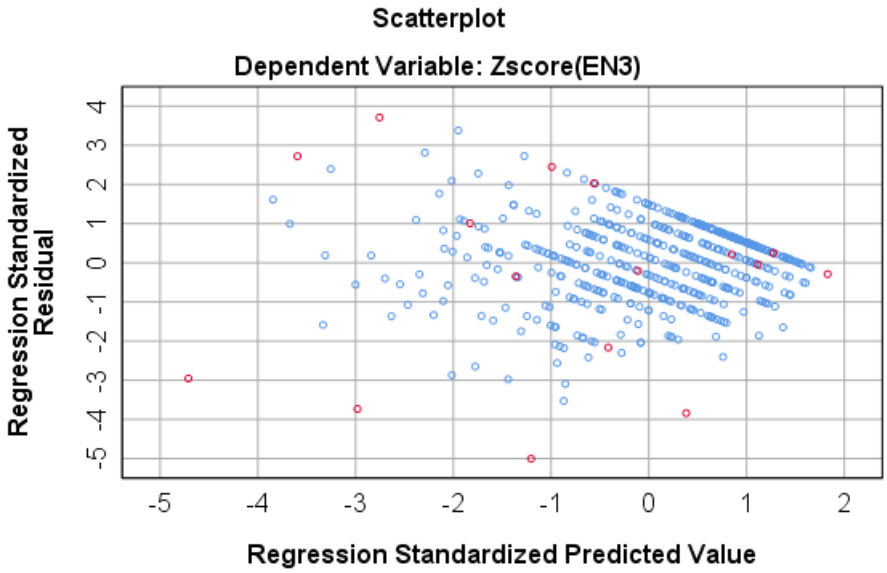


Figure 42. Residuals scatterplot for dependent variable enjoyment (ENJ). 16 cases displayed in red have been indentified as multivariate outliers (13 cases) or outliers in the solution (3 cases) and have been excluded from the analysis.

Linear model on enjoyment based only on novelty experience factors

Table 153. Sequential regression of novelty experience factors on perceived enjoyment: Standardised regression coefficient and p-values for sample without outliers $(\beta_{w/o}, p_{w/o})$, 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[54] are highlighted.

	w/o outliers			95% C. I.		
	$N = 492$			Lower	Upper	Adj.
Predictor	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)			0.724	-0.05	0.07	
Tinkering support (novelty) TIN	**0.38**	***	2E-15	0.28	0.45	**0.40**
Setting orientation (novelty) SO	0.07		0.062	0.00	0.14	0.40
Lack of problems (novelty) PRO	-0.06		0.171	-0.13	0.02	0.40
Cognitive load experiments CLE	-0.07		0.136	-0.15	0.02	0.41
Tutor support TS	**0.33**	***	7E-14	0.26	0.43	**0.47**
	$Adjusted\ R^2 = 0.47, SEE = 0.68$ $F(5,487) = 92, p = 7 \cdot 10^{-66}$					

[54] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

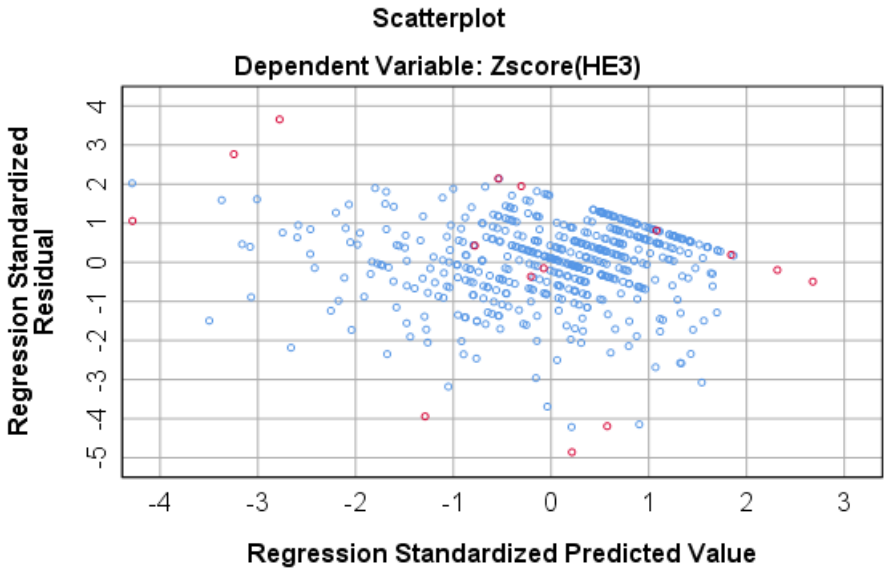


Figure 43. Residuals scatterplot for dependent variable helpfulness (HEL). 15 cases displayed in red have been indentified as multivariate outliers (13 cases) or outliers in the solution (2 additional cases) and have been excluded from the analysis.

Linear model on helpfulness based only on novelty experience factors

Table 154. Sequential regression of novelty experience factors on perceived helpfulness: Standardised regression coefficient and p-values for sample without outliers $(\beta_{w/o}, p_{w/o})$, 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[55] are highlighted.

	w/o outliers			95% C. I.		
	$N = 493$			Lower	Upper	Adj.
Predictor	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)			4E-12	-4.20	-2.39	
Tinkering support (novelty) TIN	**0.37**	***	1E-12	0.45	0.77	**0.30**
Setting orientation (novelty) SO	**0.13**	*	0.002	0.06	0.27	0.31
Lack of problems (novelty) PRO	**-0.18**	**	1E-04	-0.24	-0.08	**0.33**
Cognitive load experiments CLE	-0.01		0.856	-0.10	0.09	0.33
Tutor support TS	**0.27**	***	1E-08	0.19	0.39	**0.37**
	$Adjusted\ R^2 = 0.47, SEE = 0.68$ $F(5,487) = 92, p = 7 \cdot 10^{-66}$					

[55] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

10.9.5 SLM Situational self-beliefs: correlations matrix

Table 155. Pearson correlation r between predictors and DV situational self-beliefs SC3, significant correlations highlighted (Bonferroni correction m = 66, p < 0.0008).

		SEX	AGE	SC1	EH	EN	EO	EP	TIN	SO	PRO	CLE	TS	SC3
Gender SEX	r	1	-0.12	**-0.18**	-0.01	0.02	-0.03	-0.03	0.07	0.02	-0.04	0.07	0.05	-0.08
	p		0.008	4E-05	0.810	0.734	0.523	0.541	0.110	0.685	0.359	0.122	0.294	0.088
Age AGE	r		1	-0.01	**-0.16**	0.01	-0.01	0.13	0.05	0.14	0.10	-0.06	0.01	0.08
	p			0.816	2E-04	0.813	0.830	0.002	0.238	0.001	0.019	0.176	0.824	0.077
Dispositional self-beliefs SC1	r			1	0.03	**0.17**	0.15	0.15	**0.17**	**0.28**	0.12	**-0.23**	**0.14**	**0.38**
	p				0.464	1E-04	0.001	0.001	2E-04	2E-10	0.006	2E-07	0.001	3E-19
Experience hands-on experiments EH	r				1	0.01	0.13	-0.01	0.04	0.09	0.03	-0.03	0.04	0.07
	p					0.843	0.003	0.817	0.334	0.034	0.434	0.440	0.415	0.140
English skills (binary) EN	r					1	-0.01	0.07	0.11	0.07	0.11	-0.10	0.03	**0.18**
	p						0.859	0.120	0.010	0.104	0.014	0.022	0.434	3E-05
Experience with OSLEOs EO	r						1	0.10	0.01	0.12	-0.01	-0.06	0.09	0.14
	p							0.018	0.791	0.009	0.868	0.200	0.038	0.001
Experience with particle physics EP	r							1	0.10	**0.20**	0.08	-0.13	0.08	0.12
	p								0.022	4E-06	0.058	0.004	0.076	0.006
Tinkering support (novelty) TIN	r								1	**0.48**	**0.35**	**-0.42**	**0.57**	**0.46**
	p									2E-30	9E-16	4E-23	3E-44	8E-28
Setting orientation (novelty) SO	r									1	**0.16**	**-0.32**	**0.28**	**0.47**
	p										4E-04	3E-13	8E-11	4E-29
Lack of problems (novelty) PRO	r										1	**-0.59**	**0.26**	**0.34**
	p											1E-49	2E-09	2E-15
Cognitive load experiments CLE	r											1	**-0.32**	**-0.50**
	p												2E-13	2E-33
Tutor support TS	r												1	**0.45**
	p													3E-27

Scatterplot

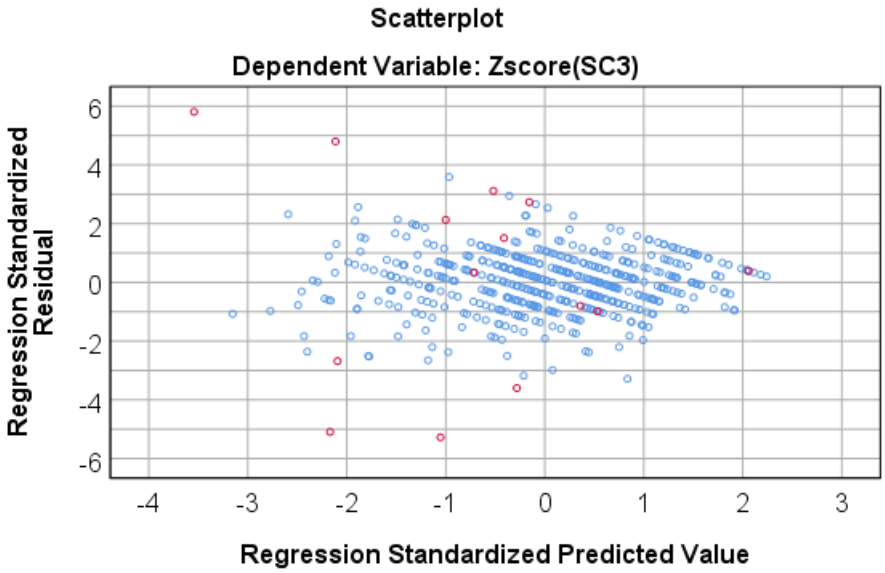


Figure 44. Residuals scatterplot for dependent variable situational self-beliefs (SC3). 14 cases displayed in red have been indentified as multivariate outliers (11 cases) or outliers in the solution (additional 3 cases) and have been excluded from the analysis.

Linear model based only on novelty experience factors

Table 156. Sequential regression of novelty experience factors on students' situational self-beliefs: Standardised regression coefficient and p-values for sample without outliers ($\beta_{w/o}$,$p_{w/o}$), 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[56] are highlighted.

	w/o outliers			95% C. I.		
	$N = 495$			Lower	Upper	Adj.
Predictor	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)			0.961	-0.06	0.06	
Tinkering support (novelty) TIN	0.09		0.053	0.00	0.18	**0.26**
Setting orientation (novelty) SO	**0.31**	***	5E-15	0.22	0.36	**0.36**
Lack of problems (novelty) PRO	0.03		0.488	-0.05	0.10	**0.40**
Cognitive load experiments CLE	**-0.35**	***	4E-15	-0.43	-0.26	**0.47**
Tutor support TS	**0.14**	**	7E-04	0.07	0.24	**0.48**
	$Adjusted\ R^2 = 0.48, SEE = 0.68$ $F(5{,}489) = 92, p = 5 \cdot 10^{-68}$					

[56] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

10.9.6 SLM Easiness: correlations matrix

Table 157. Pearson correlation r between predictors and DV easiness EAS, significant correlations highlighted (Bonferroni correction m = 55, p < 0.0009).

		SC1	EH	EN	EO	EP	TIN	SO	PRO	CLE	TS	EAS
Dispositional self-beliefs SC1	r	1	0.03	**0.17**	**0.15**	**0.15**	**0.17**	**0.28**	**0.12**	**-0.23**	**0.14**	**0.17**
	p		0.464	1E-04	7E-04	6E-04	2E-04	2E-10	6E-03	2E-07	1E-03	1E-04
Experience hands-on experiments EH	r		1	0.01	0.13	-0.01	0.04	0.09	0.03	-0.03	0.04	0.07
	p			0.843	0.003	0.817	0.334	0.034	0.434	0.440	0.415	0.106
English skills (binary) EN	r			1	-0.01	0.07	0.11	0.07	0.11	-0.10	0.03	0.08
	p				0.859	0.120	0.010	0.104	0.014	0.022	0.434	0.067
Experience with OSLEOs EO	r				1	0.10	0.01	0.12	-0.01	-0.06	0.09	0.10
	p					0.018	0.791	0.009	0.868	0.200	0.038	0.030
Experience with particle physics EP	r					1	0.10	**0.20**	0.08	-0.13	0.08	**0.17**
	p						0.022	4E-06	0.058	0.004	0.076	2E-04
Tinkering support (novelty) TIN	r						1	**0.48**	**0.35**	**-0.42**	**0.57**	**0.55**
	p							2E-30	9E-16	4E-23	3E-44	2E-41
Setting orientation (novelty) SO	r							1	**0.16**	**-0.32**	**0.28**	**0.44**
	p								4E-04	3E-13	8E-11	5E-26
Lack of problems (novelty) PRO	r								1	**-0.59**	**0.26**	**0.47**
	p									1E-49	2E-09	8E-29
Cognitive load experiments CLE	r									1	**-0.32**	**-0.61**
	p										2E-13	1E-53
Tutor support TS	r										1	**0.41**
	p											9E-22

Scatterplot

Figure 45. Residuals scatterplot for dependent variable easiness (EAS). 15 cases displayed in red have been indentified as multivariate outliers (14 cases) or outliers in the solution (1 case) and have been excluded from the analysis.

Linear model based only on novelty experience factors

Table 158. Sequential regression of novelty experience factors on perceived easiness: Standardised regression coefficient and p-values for sample without outliers $(\beta_{w/o}, p_{w/o})$, 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[57] are highlighted.

	w/o outliers $N = 493$			95% C. I.		
Predictor	$\beta_{w/o}$		$p_{w/o}$	Lower bound	Upper bound	**Adj. R^2**
(Constant)			0.706	-0.07	0.05	
Tinkering support (novelty) TIN	**0.24**	***	7E-08	0.16	0.33	**0.32**
Setting orientation (novelty) SO	**0.17**	***	2E-06	0.10	0.25	**0.37**
Lack of problems (novelty) PRO	**0.14**	**	5E-04	0.06	0.22	**0.46**
Cognitive load experiments CLE	**-0.36**	***	8E-17	-0.46	-0.29	**0.53**
Tutor support TS	0.03		0.455	-0.05	0.12	0.53
	$Adjusted\ R^2 = 0.48, SEE = 0.68$ $F(5,488) = 110, p = 2 \cdot 10^{-77}$					

[57] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

10.9.7 SLM Cognitive activation: correlations matrix

Table 159. Pearson correlation r between predictors and DV cognitive activation CA, significant correlations highlighted (Bonferroni correction m = 78, p < 0.0006).

		UN1	PH	IE1	SC1	EH	EN	EO	EP	TIN	SO	PRO	CLE	TS	CA
Pre-visit understanding score UN1	r	1	**0.25**	**0.24**	**0.29**	0.04	**0.16**	-0.01	**0.17**	0.13	0.09	**0.17**	-0.15	0.07	0.09
	p		2E-07	7E-07	1E-09	0.369	9E-04	0.897	6E-04	0.007	0.055	4E-04	0.002	0.149	0.081
Physics Grade PH1	r		1	**0.33**	**0.56**	-0.07	**0.35**	0.01	0.09	0.06	0.05	0.08	-0.11	0.02	**0.17**
	p			2E-14	1E-42	0.097	5E-16	0.767	0.056	0.146	0.272	0.082	0.012	0.618	9E-05
Dispositional Physics interest IE1	r			1	**0.65**	0.02	0.14	**0.16**	**0.15**	**0.31**	**0.40**	0.11	**-0.28**	**0.22**	**0.40**
	p				3E-63	0.724	0.002	2E-04	7E-04	7E-13	5E-21	0.018	2E-10	7E-07	2E-20
Dispositional self-beliefs SC1	r				1	0.03	**0.17**	**0.15**	**0.15**	**0.17**	**0.28**	0.12	**-0.23**	**0.14**	**0.31**
	p					0.464	1E-04	7E-04	6E-04	2E-04	2E-10	0.006	2E-07	1E-03	1E-12
Experience hands-on experiments EH	r					1	0.01	0.13	-0.01	0.04	0.09	0.03	-0.03	0.04	0.03
	p						0.843	0.003	0.817	0.334	0.034	0.434	0.440	0.415	0.561
English skills (binary) EN	r						1	-0.01	0.07	0.11	0.07	0.11	-0.10	0.03	0.14
	p							0.859	0.120	0.010	0.104	0.014	0.022	0.434	0.002
Experience with OSLEOs EO	r							1	0.10	0.01	0.12	-0.01	-0.06	0.09	0.13
	p								0.018	0.791	0.009	0.868	0.200	0.038	0.003
Experience with particle physics EP	r								1	0.10	**0.20**	0.08	-0.13	0.08	0.07
	p									0.022	4E-06	0.058	0.004	0.076	0.105
Tinkering support (novelty) TIN	r									1	**0.48**	**0.35**	**-0.42**	**0.57**	**0.60**
	p										2E-30	9E-16	4E-23	3E-44	4E-50
Setting orientation (novelty) SO	r										1	**0.16**	**-0.32**	**0.28**	**0.51**
	p											4E-04	3E-13	8E-11	1E-35
Lack of problems (novelty) PRO	r											1	**-0.59**	**0.26**	**0.21**
	p												1E-49	2E-09	2E-06
Cognitive load experiments CLE	r												1	**-0.32**	**-0.36**
	p													2E-13	1E-16
Tutor support TS	r													1	**0.48**
	p														2E-30

Scatterplot

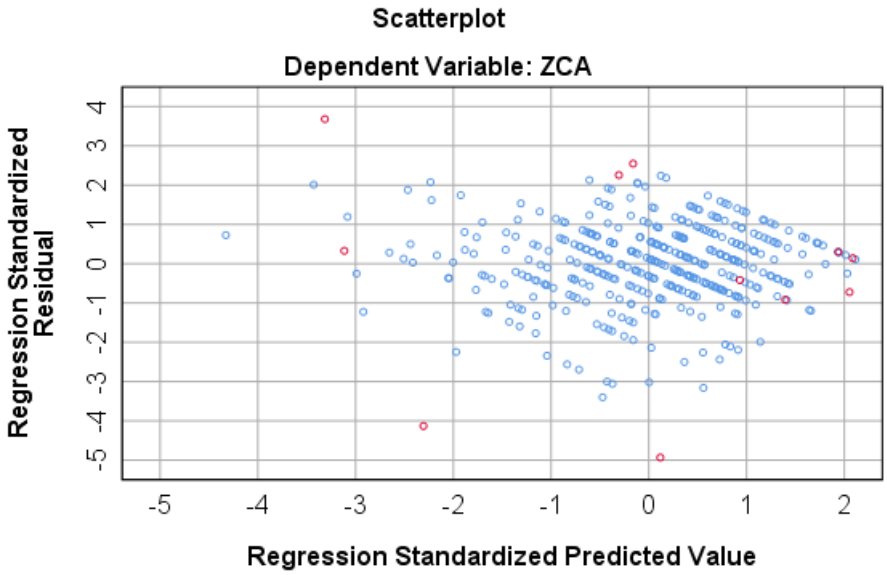


Figure 46. Residuals scatterplot for dependent variable cognitive activation (CA). 11 cases displayed in red have been indentified as multivariate outliers (9 cases) or outliers in the solution (2 cases) and have been excluded from the analysis.

Linear model based only on novelty experience factors

Table 160. Sequential regression of novelty experience factors on cognitive activation: Standardised regression coefficient and p-values for sample without outliers ($\beta_{w/o}, p_{w/o}$), 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[58] are highlighted.

	w/o outliers			95% C. I.		
	$N = 495$			Lower	Upper	Adj.
Predictor	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)			0.979	-0.06	0.06	
Tinkering support (novelty) TIN	**0.30**	***	1E-10	0.21	0.39	**0.35**
Setting orientation (novelty) SO	**0.27**	***	7E-12	0.19	0.33	**0.42**
Lack of problems (novelty) PRO	-0.07		0.086	-0.15	0.01	0.41
Cognitive load experiments CLE	-0.13		0.004	-0.22	-0.04	**0.43**
Tutor support TS	**0.22**	***	1E-07	0.14	0.30	**0.46**
	$Adjusted\ R^2 = 0.46, SEE = 0.68$ $F(5{,}492) = 85, p = 5 \cdot 10^{-64}$					

[58] The Bonferroni correction reduces the required p-values with $m = 12$ to
* $p < 0.004$, ** $p < 0.0008$ and *** $p < 0.00008$

10.9.8 SLM Conceptual understanding post: correlations matrix

Table 161. Pearson correlation r between predictors and DV concept. understanding UN3, significant correlations highlighted (Bonferroni correction m = 78, p < 0.0006).

		UN1	PH1	EH	EN	EO	EP	TIN	SO	PRO	CLE	TS	UN3
Pre-visit understanding score UN1	r	1	**0.25**	0.04	**0.16**	-0.01	**0.17**	0.13	0.09	**0.17**	-0.15	0.07	**0.55**
	p		2E-07	0.369	9E-04	0.897	6E-04	0.007	0.055	4E-04	0.002	0.149	8E-31
Physics Grade PH1	r		1	-0.07	**0.35**	0.01	0.09	0.06	0.05	0.08	-0.11	0.02	**0.33**
	p			0.097	5E-16	0.767	0.056	0.146	0.272	0.082	0.012	0.618	4E-13
Experience hands-on experiments EH	r			1	0.01	0.13	-0.01	0.04	0.09	0.03	-0.03	0.04	-0.09
	p				0.843	0.003	0.817	0.334	0.034	0.434	0.440	0.415	0.057
English skills (binary) EN	r				1	-0.01	0.07	0.11	0.07	0.11	-0.10	0.03	**0.23**
	p					0.859	0.120	0.010	0.104	0.014	0.022	0.434	5E-07
Experience with OSLEOs EO	r					1	0.10	0.01	0.12	-0.01	-0.06	0.09	0.02
	p						0.018	0.791	0.009	0.868	0.200	0.038	0.694
Experience with particle physics EP	r						1	0.10	**0.20**	0.08	-0.13	0.08	**0.20**
	p							0.022	4E-06	0.058	0.004	0.076	1E-05
Tinkering support (novelty) TIN	r							1	**0.48**	**0.35**	**-0.42**	**0.57**	**0.20**
	p								2E-30	9E-16	4E-23	3E-44	2E-05
Setting orientation (novelty) SO	r								1	**0.16**	**-0.32**	**0.28**	0.08
	p									4E-04	3E-13	8E-11	0.107
Lack of problems (novelty) PRO	r									1	**-0.59**	**0.26**	**0.28**
	p										1E-49	2E-09	1E-09
Cognitive load experiments CLE	r										1	**-0.32**	**-0.25**
	p											2E-13	6E-08
Tutor support TS	r											1	0.14
	p												0.003

Scatterplot

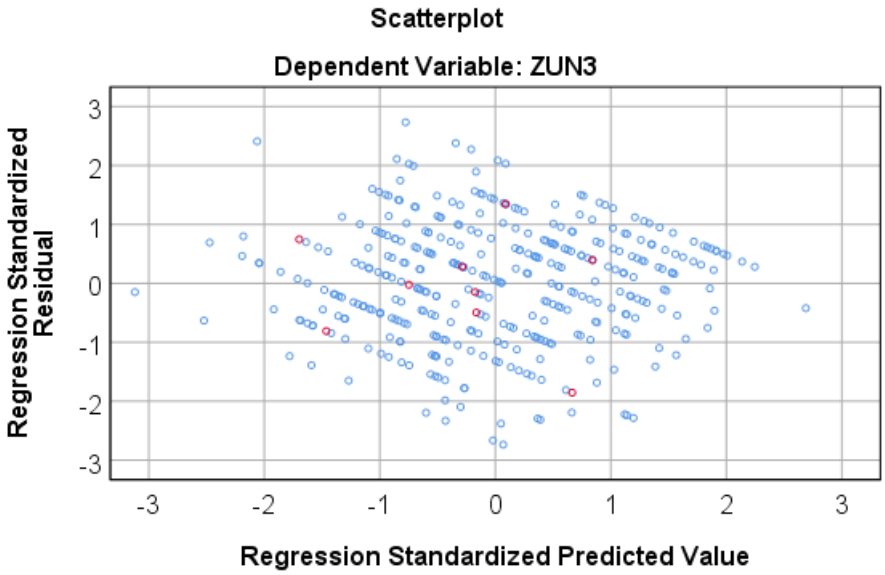


Figure 47. Residuals scatterplot for dependent variable understanding post (UN3). 10 cases displayed in red have been indentified as multivariate outliers and have been excluded from the analysis.

Linear model based only on novelty experience factors

Table 162. Sequential regression of novelty experience factors on students' conceptual understanding: Standardised regression coefficient and p-values for sample without outliers $(\beta_{w/o}, p_{w/o})$, 95% confidence interval for standardised regression coefficients, and adjusted R^2 per regression level. Significant predictors[59] are highlighted.

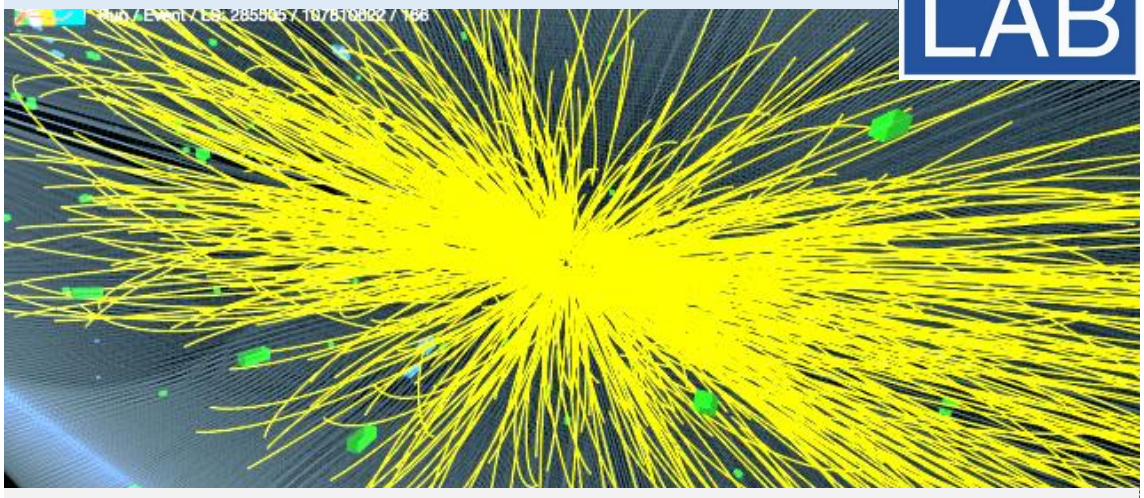
	w/o outliers			95% C. I.		
	$N = 365$			Lower	Upper	Adj.
Predictor	$\beta_{w/o}$		$p_{w/o}$	bound	bound	R^2
(Constant)			0.810	-0.10	0.08	
Tinkering support (novelty) TIN	0.12		0.074	-0.01	0.26	**0.04**
Setting orientation (novelty) SO	-0.04		0.424	-0.15	0.06	0.04
Lack of problems (novelty) PRO	**0.18**	*	0.002	0.07	0.30	**0.08**
Cognitive load experiments CLE	-0.10		0.122	-0.23	0.03	0.09
Tutor support TS	0.00		0.981	-0.13	0.13	0.08
	$Adjusted\ R^2 = 0.09, SEE = 0.96$ $F(5,438) = 9, p = 2 \cdot 10^{-8}$					

[59] The Bonferroni correction reduces the required p-values with $m = 11$ to
* $p < 0.004$, ** $p < 0.0009$ and *** $p < 0.00009$

10.10 Final questionnaire (MTF items including item easiness)

CERN S'Cool LAB

Questions about Radioactivity

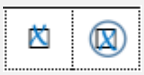


Dear student,

we, the S'Cool LAB team, want to understand how students learn about particle physics. In this questionnaire, we would like to find out, what your conceptions of radioactivity are.

Your answers will **remain anonymous** and **this is not a test**!

Many of the questions might be difficult for you, because some of the topics are not taught at school. If you don't know the answer, follow you intuition!

- Please choose for every statement, whether you think, it is TRUE or FALSE.
- Approximately 30% of the statements will be true.
- Please do not talk to your neighbours while filling out the questionnaire.
- If you don't know a word in English, just write the word in your own language.
- If you want to change your answer, simply put a new cross in the right place and draw a circle around the new cross 

Many thanks in advance for your help!

Contact: CERN S'Cool LAB cern.ch/s-cool-lab

Example question

Reminder atom: orbital and nucleus

Atoms consist of protons, neutrons, and electrons. Neutrons and protons can be found in the nucleus space, electrons can be found in the orbital space around the nucleus space. Most physics phenomena can be traced back to interactions of these components with other particles.

Note: nuclei refers to the plural of nucleus

Which particles form the atomic nucleus?	TRUE ✓	FALSE ✗
Electrons and protons.	☐	☐
Protons and neutrons.	☐	☐
Neutrons and electrons.	☐	☐
Photons and neutrons.	☐	☐

Radioactive sources & unstable atoms I

	A There are different types of radioactive decays, every type is characterised by a certain process in the atom.	TRUE ✓	FALSE ✗
A1	During one type of radioactive decay, an electron is emitted from the orbital of an atom.	□	56%
A2	During one type of radioactive decay, an electron is emitted from the nucleus of an atom.	45%	□
A3	During one type of radioactive decay, an electron drops from a higher to a lower energy level releasing energy.	□	48%
A4	During one type of radioactive decay, some combination of protons and neutrons are emitted from the nucleus.	68%	□
A5	During one type of radioactive decay, the nucleus drops from a higher to a lower energy level releasing energy.	28%	□
A6	During one type of radioactive decay, the atom slowly disappears over time until nothing is left.	□	85%
A7	During one type of radioactive decay, the nucleus slowly disappears over time until nothing is left.	□	83%

	B What characterises radioactive atoms?	TRUE ✓	FALSE ✗
B1	Some atoms are radioactive because they consist of an unbalanced number of electrons and protons.	□	73%
B2	Some atoms are radioactive because they consist of an unbalanced number of neutrons and protons.	69%	□
B3	Some atoms are radioactive because they consist of an unbalanced number of electrons and neutrons.	□	92%
B4	Some atoms are radioactive because they are in an excited state with too much energy.	67%	□
B5	Some atoms are radioactive because they have existed too long.	□	87%
B6	Some atoms are radioactive because they have been in contact with special chemicals (for example strong acids).	□	67%
B7	All atoms are radioactive because everything is a little bit radioactive.	□	61%

Radioactive sources II

F	What is the half-life?	TRUE ✓	FALSE ✗
F1	The half-life is the time in which the nucleus of a radioactive atom loses half of its radiation.	□	70%
F2	The half-life is the time in which a radioactive substance is dangerous.	□	97%
F3	The half-life is the time in which a radioactive substance loses half of its mass.	□	55%
F4	The half-life is the time in which the nuclei of a radioactive substance have been split into two parts of approximately the same size.	□	89%
F5	The half-life is the time in which half of the radioactive atoms of a substance decayed by emitting radiation.	79%	□
F6	The half-life is the time in which the volume of a radioactive substance has been reduced to half its amount.	□	70%

G	**A substance (mass 100 g, volume 80 cm^3) is made of 1000 radioactive atoms with a half-life of 1 hour.** How many radioactive atoms are left after 1 hour? How many atoms in total are left after 1 hour?	TRUE ✓	FALSE ✗
G1	After 1 hour, the total number of atoms is 1500, because half of the radioactive atoms decayed into 2 stable atoms.	□	83%
G2	After 1 hour, the volume of the substance is reduced to approximately 40$cm^{3.}$	□	72%
G3	After 1 hour, the mass of the substance is reduced to approximately 50g.	□	60%
G5	After 1 hour, approximately 500 of the radioactive atoms transformed into other atoms.	68%	□
G6	After 1 hour, there will still be 1000 radioactive atoms, but their radiation will be weaker.	□	79%
G7	After 1 hour, the total number of atoms will be 500, because 500 radioactive atoms decayed and disappeared.	□	80%

Radioactive sources II

H	Which properties of a substance determine the half-life of the radioactive atoms in this substance?	TRUE ✓	FALSE ✗
H1	The half-life depends on the temperature of the substance.	□	78%
H2	The half-life depends on the pressure of the substance.	□	80%
H3	The half-life depends on the density of the substance.	□	66%
H4	The half-life depends on the chemicals that were in contact with the substance.	□	77%
H5	The half-life depends on the properties of the nuclei (number of protons and neutrons) in the substance.	85%	□
H6	The half-life depends on the age of the substance.	□	79%
H7	The half-life depends on the properties of the orbital (number of electrons) in the substance.	□	62%

Ionising radiation I

I	What is ionisation?	TRUE ✓	FALSE ✗
I1	During ionisation, electrically neutral atoms lose or gain electrons in the orbital, therefore they become electrically charged.	88%	□
I2	During ionisation, stable atoms become radioactive.	□	84%
I3	During ionisation, electrons in the orbital are excited to higher energy levels, but they are not able to leave the atom.	□	85%
I4	During ionisation, the nucleus is excited to higher energy levels.	□	86%
I6	During ionisation, electrically neutral atoms lose or gain protons, therefore they become electrically charged.	□	84%

J	There are different types of ionising radiation.	TRUE ✓	FALSE ✗
J1	Some types of ionising radiation can be described with a particle model.	80%	□
J2	Some types of ionising radiation can be described with a wave model.	70%	□
J3	Some types of ionising radiation consist of radioactive particles, which can travel through air and will decay later.	□	57%
J4	Some types of ionising radiation consist of pure energy, and can't be described by particle models or wave models.	□	77%
J5	Some types of ionising radiation are electrically charged.	63%	□
J8	Some types of ionising radiation can be dangerous for humans.	□	90%

Ionising radiation II

K	Can the following types of radiation ionise an atom?	YES ✓	NO ✗
K8	X-rays	76%	□
K2	Radio waves	□	81%
K3	Mobile phone radiation	□	88%
K5	Cosmic radiation	71%	□
K16	Wi-Fi radiation	□	94%
K7	Alpha radiation (Helium nuclei)	71%	□
K6	Beta minus radiation (electrons)	80%	□
K4	Gamma radiation (high energy photons)	87%	□
K9	Micro waves	□	68%
K10	Laser light from a red laser pointer	□	83%
K11	Infrared radiation (IR)	□	70%
K12	Ultraviolet Light (UV light)	63%	□
K13	White light from a lamp	□	90%
K14	Ultrasound used in hospitals	□	20%
K15	Radiation from batteries	□	76%

Interaction of radiation with matter I

L	How can you turn stable atoms into radioactive atoms?	TRUE ✓	FALSE ✗
L1	Stable atoms can become radioactive if you bring them in contact with special chemicals.	□	64%
L2	Stable atoms can become radioactive if you expose them to neutrons.	40%	□
L3	Stable atoms can become radioactive if you expose them to high energy protons.	58%	□
L4	Stable atoms can become radioactive if you heat them up.	□	91%
L5	Stable atoms can become radioactive if you place them in an X-ray machine and irradiate them for 1 hour.	□	49%
L6	Stable atoms can become radioactive if you place them in a 10 cm distance to a radioactive source for a while.	□	49%

M	Doctors forgot to switch off the X-ray machine in a hospital. Inside the X-ray machine, they forgot a ring made of copper. What happens to the copper when is it exposed to X-rays for a 1 hour?	TRUE ✓	FALSE ✗
M1	After 1 hour, the copper will get very warm, it will slowly start melting.	□	66%
M2	After 1 hour, the copper will start to smell.	□	87%
M4	After 1 hour, the radiation will accumulate in front of the copper block.	□	78%
M5	After 1 hour, the radiation will accumulate inside the copper block.	□	54%
M6	Most of the radiation will be reflected by the smooth surface.	□	82%
M7	After 1 hour, some of the electrons in the copper atoms will be excited to higher energy levels.	75%	□
M8	After 1 hour, some of the chemical bonds between the copper atoms will break up, the structure of the metal might change a bit.	53%	□
M9	After 1 hour, the copper will start to glow in the dark.	□	91%
M10	After 1 hour, the copper will be radioactive.	□	60%

Interaction of radiation with matter III

N	Doctors forgot to switch off the X-ray machine in a hospital. Inside the X-ray machine, they forgot a patient. What happens to the body of the patient when is it exposed to X-rays for a 1 hour?	TRUE ✓	FALSE ✗
N1	After 1 hour, the body will get very warm, the water in the body will boil.	□	60%
N2	After 1 hour, the patient will smell differently.	□	83%
N4	After 1 hour, the radiation will accumulate in front of the body.	□	84%
N5	After 1 hour, the radiation will accumulate inside the body.	□	38%
N6	Most of the radiation will be reflected by the skin and the bones.	□	80%
N7	After 1 hour, the patient will start to glow in the dark.	□	95%
N8	After 1 hour, the DNA helix in some cells will be destroyed.	85%	□
N9	After 1 hour, the person will get poisoned.	□	67%
N10	After 1 hour, some cells that have been exposed to radiation will die.	89%	□
N11	After 1 hour, the person will lose their hair.	□	73%
N12	After 1 hour, the person will be radioactive.	□	72%
N13	After 1 hour, the some cells that have been exposed to radiation will start to mutate.	77%	□

Interaction of radiation with matter III

O	How dangerous radiation is for humans depends on …	TRUE ✓	FALSE ✗
O1	… the energy of the radiation.	77%	□
O2	… the type of radiation.	92%	□
O3	… the intensity of radiation.	88%	□
O4	… the distance to the radiation source.	88%	□
O5	… the exposure time.	94%	□
O6	… the part of the body which absorbs the radiation.	44%	□

X-rays: Risks for cleaning staff in the radiology department?

P	Is it dangerous for the cleaning staff of the radiology department of a hospital to clean a room with an X-ray machine?	TRUE ✓	FALSE ✗
P1	**It is dangerous**, because parts of the device are exposed to X-rays and therefore become radioactive themselves.	□	80%
P2	**It is dangerous**, because of X-ray machines are always a source of radiation, even if they are out of service.	□	77%
P3	**It is NOT dangerous,** because the X-ray units are out of service during the cleaning.	59%	□
P4	**It is NOT dangerous**, provided that the entire air of the treatment rooms has been exchanged before the cleaning.	□	78%

X-rays: Danger for visitors?

Q	After a sports accident the foot of a patient is often X-rayed to check whether there are bone fractures. Is it dangerous for the patient's family to be near the patient immediately after the X-ray scan?	TRUE ✓	FALSE ✗
Q1	**It is dangerous**, because the foot has been exposed to X-rays and therefore becomes radioactive itself.	□	94%
Q2	**It is dangerous**, because the foot accumulates ionizing radiation, and therefore starts to radiate itself.	□	89%
Q3	**It is NOT dangerous**, because X-ray machines do not transfer radioactive atoms to the foot.	75%	□
Q4	**It is NOT dangerous**, but you have to wait a bit before visiting the patient to make sure that the radiation is gone.	□	73%

X-rays: A watch in an X-ray machine

R	A watch is X-rayed for 10 min inside an X-ray machine. Is it still okay to wear the watch afterwards?	TRUE ✓	FALSE ✗
R1	**It is OK to wear the watch**. To become radioactive itself, the watch needs to be exposed to X-rays much longer.	□	68%
R2	**It is OK to wear the watch**. As long as no radioactive atoms are transferred to the watch, the watch does not emit radiation.	54%	□
R3	**You should NOT wear the watch**, because ionizing radiation accumulates in the watch, therefore it starts to radiate itself.	□	77%
R4	**You should NOT wear the watch**, because the watch has been exposed to X-rays and therefore becomes radioactive itself.	□	85%

Some information about you

Z1	How old are you?	_____ years	
Z2	You are …	☐ female	☐ male
Z3	What grade are you in? / How many years have you been going to school? *(e.g., 11 years → grade 11)*		
Z4	What is your native language?		
Z6	In what country is your school?		
	How good are you in Physics? *(You don't need to think about your last grades, just think about how you feel about your performance in Physics lessons in general)*	◯ very good ◯ good ◯ satisfactory / average ◯ bad ◯ very bad	
	How difficult was it for you to understand the English words and sentences in this questionnaire? ◯ I understood all the words and sentences without problems. ◯ I did not know all the words, but I understood the meaning of the questions. ◯ I did not understand all the questions because I had problems with some of the words. ◯ I did not understand most of the questions.		
Z14	If you are currently taking physics: How many minutes per week do you spend in physics classes?	_____ minutes	

How do you like your physics classes?

	Please rate the following statements.	completely disagree	disagree	disagree a little	agree a little	agree	completely agree
IE1	I spend time on physics in my free time – even beyond homework.	☐	☐	☐	☐	☐	☐
IE3	I enjoy solving physics problems.	☐	☐	☐	☐	☐	☐
IE4	I discuss topics from my physics classes with my friends or my family.	☐	☐	☐	☐	☐	☐
IE7	When I am working on physics problems, sometimes, I do not notice time passing.	☐	☐	☐	☐	☐	☐
IE8	I like our physics classes.	☐	☐	☐	☐	☐	☐
IE9	I would like to spend more time in school on physics.	☐	☐	☐	☐	☐	☐

10.11 Items that were excluded because of ambiguity

Table 163. Overview of item stems, and individual items that were excluded because of ambiguity.

Item stem		Nr. of items in total	Nr. of excluded items
A	**There are different types of radioactive decays, every type is characterised by a certain process in the atom.**	**8**	**1**
A8	During one type of radiative decay, the atom breaks up into 2 stable atoms.		
B	**What characterises radioactive atoms?**	**7**	**0**
F	**What is the half-life?**	**6**	**0**
G	**A substance (mass 100 g, volume 80 cm3) is made of 1000 radioactive atoms with a half-life of 1 hour. How many radioactive atoms are left after 1 hour? How many atoms in total are left after 1 hour?**	**8**	**2**
G4	After 1 hour, the number of radioactive atoms is reduced to approximately 500.		
G8	After 1 hour, there will still be 1000 atoms in total, but it is not possible to say how many of them will be radioactive.		
H	**Which properties of a substance determine the half-life of the radioactive atoms in this substance?**	**7**	**0**
I	**What is ionisation?**	**6**	**1**
I5	During ionisation, ions emit radiation.		
J	**There are different types of ionising radiation.**	**8**	**2**
J6	Some types of ionising radiation are emitted by ions.		
J7	Some types of ionising radiation make stable atoms radioactive.		
K	**Can the following types of radiation ionise an atom?**	**16**	**1**
K1	Heat		
L	**How can you turn stable atoms into radioactive atoms?**	**7**	**1**
L7	Stable atoms will become radioactive if you bring them in direct contact with a radioactive source for a while.		
M	**Doctors forgot to switch off the X-ray machine in a hospital. Inside the X-ray machine, they forgot a ring made of copper. What happens to the copper when is it exposed to X-rays for a 1 hour?**	**10**	**1**
M3	After 1 hour, the colour of the copper will change.		
N	**Doctors forgot to switch off the X-ray machine in a hospital. Inside the X-ray machine, they forgot a patient. What happens to the body of the patient when is it exposed to X-rays for a 1 hour?**	**13**	**1**
N3	After 1 hour, the colour of the skin will change.		
O	**How dangerous radiation is for humans depends on …**	**7**	**1**
O7	… whether the material is a living organism or not.		
	SUM	**103**	**11**

10.12 Scale properties of the extracted factors

10.12.1 Factor 1 (16 items)

Table 164. Item label and description, corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}).

Label	Item short description	r_{tj}	α_{tj}
A4	Decay = p^+ and n^0 are emitted from the nucleus	0.36	0.69
B2	Radioactive = unbalanced number of n^0 and p^+	0.37	0.69
F5	T1/2 = half the radioactive atoms decayed	0.20	0.71
G5	T1/2 = 500 of 1000 radioact. atoms transformed	0.23	0.71
H5	T1/2 depends on the properties of the nucleus	0.39	0.69
I1	Ionisation = atoms loose or gain electrons	0.21	0.70
J1	Radiation can be described by particle model	0.34	0.69
J2	Radiation can be described by wave model	0.39	0.69
J5	Radiation can be electrically charged	0.36	0.69
L2	Neutron exposure can make atoms radioactive	0.24	0.70
M7	Copper electrons are excited when X-rayed 1h	0.17	0.71
N8	DNA is destroyed in some cells when X-rayed 1h	0.41	0.68
O1	Radiation danger depends on energy	0.29	0.70
O3	Radiation danger depends on the intensity	0.28	0.70
O4	Radiation danger depends on the distance	0.26	0.70
O6	Radiation danger depends on exposed body part	0.34	0.69
	Cronbach's α_C		**0.71**

10.12.2 Factor 2 (10 items)

Table 165. Item label and description, corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}).

Label	Item short description	r_{tj}	α_{tj}
A1	Decay = an electron is emitted from the orbital	0.51	0.65
A2	Decay = an electron is emitted from the nucleus	0.48	0.65
A3	Decay = an electron drops to lower energy level	0.34	0.68
A5	Decay = the nucleus drops to lower energy level	0.18	0.71
B1	Radioactive = unbalanced number of e- and p+	0.35	0.68
B6	Radioactive = atoms were in contact with chemicals	0.37	0.68
F4	T1/2 = time in which the nucleus splits into 2 parts	0.29	0.69
H4	T1/2 depends on previous contact to chemicals	0.35	0.68
H7	T1/2 on the properties of the orbital	0.30	0.69
L1	Contact to chemicals can make atoms radioactive	0.42	0.67
	Cronbach's α_C		0.70

10.12.3 Factor 3 (6 items)

Table 166. Item label and description, corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}).

Label	Item short description	r_{tj}	α_{tj}
A6	Decay = the atom slowly disappears	0.24	0.70
F3	T1/2 = radioactive substance loses half its mass	0.56	0.61
F6	T1/2 = radioactive substance loses half its volume	0.46	0.64
G2	T1/2 = 40cm3 left of 80 cm3 radioactive substance	0.56	0.61
G3	T1/2 = 50g left of 100g radioactive substance	0.49	0.63
G7	T1/2 = 500 of 1000 atoms disappeared	0.23	0.71
	Cronbach's α_C		0.70

10.12.4 Factor 4 (6 items)

Table 167. Item label and description, corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}).

Label	Item short description	r_{tj}	α_{tj}
J3	Radiation can consist of radioactive particles	0.12	0.60
L6	Source (10cm away) can make atoms radioactive	0.22	0.56
M5	X-radiation accumulates inside the copper	0.35	0.50
M10	Copper becomes radioactive when X-rayed 1h	0.42	0.47
N5	X-radiation accumulates inside the body	0.39	0.48
N12	Person becomes radioactive when X-rayed 1h	0.35	0.50
	Cronbach's α_C		0.57

10.12.5 Factor 5 (10 items)

Table 168. Item label and description, corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}).

Label	Item short description	r_{tj}	α_{tj}
A7	Decay = the nucleus slowly disappears	0.23	0.48
H1	T1/2 depends on the temperature of a substance	0.29	0.46
H2	T1/2 depends on the pressure of a substance	0.37	0.43
H3	T1/2 depends on the density of a substance	0.20	0.49
H6	T1/2 depends on the age of a substance	0.18	0.50
M4	X-radiation accumulates in front of copper	0.17	0.50
N4	X-radiation accumulates in front of the body	0.20	0.49
N6	X-radiation is reflected by the bones in the body	0.22	0.48
N10	Some cells will die when X-rayed for 1 hour	0.12	0.51
N13	Some cells will start to mutate when X-rayed for 1 hour	0.18	0.49
	Cronbach's α_C		0.51

10.12.6 Factor 6 (6 items)

Table 169. Item label and description, corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}).

Label	Item short description	r_{tj}	α_{tj}
M1	Copper gets warm when X-rayed 1h	0.26	0.52
M2	Copper smells when X-rayed 1h	0.26	0.52
N1	Body gets warm when X-rayed 1h	0.38	0.46
N2	Body smells when X-rayed 1h	0.34	0.49
N9	Person gets poisoned when X-rayed 1h	0.28	0.51
N11	Person will lose hair when X-rayed 1h	0.26	0.52
	Cronbach's α_C		0.55

10.12.7 Factor 7 (4 items)

Table 170. Item label and description, corrected item-scale correlation (r_{tj}), and Cronbach's alpha when omitting item j (α_{tj}).

Label	Item short description	r_{tj}	α_{tj}
B7	Everything is a little bit radioactive	0.22	0.21
G1	T1/2 = half the atoms decayed into 2 stable atoms	0.09	0.36
I3	Ionisation = electrons excited to higher energy level	0.20	0.25
J4	Radiation can consist of pure energy	0.19	0.25
	Cronbach's α_C		0.33

10.12.8 Items that were not interpreted within the seven-factor solution

Table 171. Items that could not be interpreted within the 7-factor solution. This table lists the Item label, true/false T/F, item description, assigned category (Cat), factor loadings for the 7-factor solution f1-f7, and item communalities h^2 based on the rotated factor solution (Varimax, Kaiser normalization). Factor loadings |a| < 0.32 are omitted.

	TF	Item description	Cat.	f1	f2	f3	f4	f5	f6	f7	h^2
B4	T	Radioactive = excited state with too much energy	S-mi								0.2
B5	F	Radioactive = atoms already existed for too long	S-mi								0.1
F1	F	T1/2 = the nucleus loses half of its radiation	S-mi							-0.44	0.3
G6	F	T1/2 = same number of atoms but weaker radiation	S-ma							-0.67	0.7
I2	F	Ionisation = stable atoms become radioactive	A-mi								0.2
I4	F	Ionisation = nucleus excited to higher energy level	A-mi	0.32			0.34	0.39	0.34		0.5
I6	F	Ionisation = atoms loose or gain protons	A-mi								0.2
L3	T	Proton exposure can make atoms radioactive	A-mi				-0.44				0.3
L5	F	X-radiation can make atoms radioactive	A-mi			0.38					0.4
M6	F	X-radiation is reflected by the copper surface	A-ma								0.2
M8	T	Chemical bonds break when X-rayed 1h	A-mi								0.3

10.13 Overview of electronic appendix

The electronic appendix includes the following files:

A) Student instructions of the three hands-on workshops
    a. Electron tube workshop (worksheet, pdf)
    b. Cloud Chamber workshop (Do-it-yourself manual, pdf)
    c. X-ray workshop (worksheet, pdf)
B) Processed student data of the S'Cool LAB evaluation main study
C) Processed student data of the students' conceptions study

All these files are also available online at:

https://cernbox.cern.ch/index.php/s/Rv2dBV7nJ9uxKx3

11 References

The ATLAS Collaboration (2008). The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation, 3*, S08003.

Abrahams, I., & Millar, R. (2008). Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education, 30*(14), 1945.

Acar Sesen, B., & Ince, E. (2010). Internet as a Source of Misconception. *Turkish Online Journal of Educational Technology-TOJET, 9*(4), 94.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*(3), 545.

Ainley, M., Hillman, K., & Hidi, S. (2002). Gender and interest processes in response to literary texts: Situational and individual interest. *Learning and Instruction, 12*(4), 411.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716.

Alexander, P. A., & Grossnickle, E. M. (2016). Positioning interest and curiosity within a model of academic development. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of Motivation at School* (2nd Edition): Routledge.

Allen, S. (2004). Designs for learning: studying science museum exhibits that do more than entertain. *Science Education, 88*(S1).

Alsop, S. (2001). Living with and learning about radioactivity: A comparative conceptual study. *International Journal of Science Education, 23*(3), 263.

Anderson, C. D., & Neddermeyer, S. H. (1936). Cloud chamber observations of cosmic rays at 4300 meters elevation and near sea-level. *Physical Review, 50*(4), 263.

Anderson, D., & Ellenbogen, K. M. (2012). Learning science in informal contexts–Epistemological perspectives and paradigms. In *Second International Handbook of Science Education* (pp. 1179): Springer.

Anderson, D., & Lucas, K. B. (1997). The effectiveness of orienting students to the physical features of a science museum prior to visitation. *Research in Science Education, 27*(4), 485.

Andre, T., & Windschitl, M. (2003). Interest, epistemological belief, and intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional Conceptual Change:* Erlbaum.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). 'Doing' science versus 'being' a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education, 94*(4), 617.

Arnone, M. P., Small, R. V., Chauncey, S. A., & McKenna, H. P. (2011). Curiosity, interest and engagement in technology-pervasive learning environments: a new research agenda. *Educational Technology Research and Development, 59*(2), 181.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403.

Atkinson, E. P. (1990). Learning scientific knowledge in the student laboratory. In E. Hegarty-Hazel (Ed.), *The Student Laboratory and the Science Curriculum*. London: New York: Routledge.

Ausubel, D. P. (1968). *Educational psychology. A cognitive view*. New York: Holt, Rinehart and Winston.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47): Elsevier.

Bagno, E., & Eylon, B.-S. (1997). From problem solving to a knowledge structure: An example from the domain of electromagnetism. *American Journal of Physics, 65*(8), 726.

Baker, F. B. (1965). An investigation of the sampling distributions of item discrimination indices. *Psychometrika, 30*(2), 165.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning Analytics* (pp. 61): Springer.

Bamberger, Y., & Tal, T. (2007). Learning in a personal context: Levels of choice in a free choice learning environment in science and natural history museums. *Science Education, 91*(1), 75.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review, 84*(2), 191.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133.

Bax, S. (2016). Text Inspector. Retrieved on 14/03/2020 from https://textinspector.com

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289.

Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal, 21*(4), 755.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York, NY, US: McGraw-Hill Book Company.

Berlyne, D. E. (1978). Curiosity and learning. *Motivation and Emotion, 2*(2), 97.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In *Assessment and Teaching of 21st Century Skills* (pp. 17): Springer.

Boiko, A., Nistor, A., Kudenko, I., & Gras-Velazquez, A. (2019). *The attractiveness of Science, Technology, Engineering and Mathematics subjects. Results from five countries*. Brussels: European Schoolnet.

Börlin, J. (2012). *Das Experiment als Lerngelegenheit: vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität* (Vol. 132): Logos Verlag.

Boscolo, P., Ariasi, N., Del Favero, L., & Ballarin, C. (2011). Interest in an expository text: How does it flow from reading to writing? *Learning and Instruction, 21*(3), 467.

Boyes, E., & Stanisstreet, M. (1994). Children's Ideas about Radioactivity and Radiation: sources, mode of travel, uses and dangers. *Research in Science & Technological Education, 12*(2), 145.

Brandt, A. (2005). *Förderung von Motivation und Interesse durch außerschulische Experimentierlabors*: Cuvillier Verlag.

Brassil, C. E., & Couch, B. A. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. *International Journal of STEM Education, 6*(1), 16.

Braund, M., & Reiss, M. (2006). Towards a more authentic science curriculum: The contribution of out-of-school learning. *International Journal of Science Education, 28*(12), 1373.

Brell, C. (2008). *Lernmedien und Lernerfolg - reale und virtuelle Materialien im Physikunterricht: empirische Untersuchungen in achten Klassen an Gymnasien (Laborstudie) zum Computereinsatz mit Simulation und IBE*: Logos Verlag.

Brod, G., & Breitwieser, J. (2019). Lighting the wick in the candle of learning: generating a prediction stimulates curiosity. *NPJ Science of Learning, 4*(1), 1.

Brod, G., Breitwieser, J., Hasselhorn, M., & Bunge, S. A. (2019). Being proven wrong elicits learning in children - but only in those with higher executive function skills. *Developmental Science*, e12916.

Burge, B., Lenkeit, J., & Sizmur, J. (2015). *PISA in practice: Cognitive activation in maths*: National Foundation for Educational Research.

Burkam, D. T., Lee, V. E., & Smerdon, B. A. (1997). Gender and science learning early in high school: Subject matter and laboratory experiences. *American Educational Research Journal, 34*(2), 297.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261.

Buse, M. (2018). *Bilinguale englische experimentelle Lehr-Lernarrangements im Fach Biologie-Konzeption, Durchführung und Evaluation der kognitiven und affektiven Wirksamkeit.* (PhD), Bergische Universität Wuppertal, Wuppertal.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380.

Carlone, H. B. (2004). The cultural production of science in reform ‐ based physics: Girls' access, participation, and resistance. *Journal of Research in Science Teaching, 41*(4), 392.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245.

Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research, 12*(1), 43.

Chabay, R., & Sherwood, B. (1997a). *Brief Electricity and Magnetism Assessment (BEMA)*. Retrieved on 14/03/2020 from PhysPort https://www.physport.org/

Chabay, R., & Sherwood, B. (1997b). Qualitative understanding and retention. *AAPT Announcer, 27*(2), 96.

Chasteen, S. V., & Pollock, S. J. (2009). Tapping into juniors' understanding of E&M: The colorado upper-division electrostatics (CUE) diagnostic. *Physics Education Research Conference (PERC) Proceedings 2009*, 109.

Chi, B., Dorph, R., & Reisman, L. (2015). Evidence & impact: Museum-managed STEM programs in out-of-school settings. *National Research Council Committee on Out-of-School Time STEM. Washington, DC: National Research Council*.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219.

Clement, P., & Fisseux, C. (1999). Opacity of radiography, perplexity of teachers and pupils in Primary School. In *Research in Science Education in Europe*: Springer.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd. In: Hillsdale, NJ: Erlbaum.

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34*(3), 315.

Colclough, N. D. (2007). *Trainee teachers and ionising radiation: understandings, attitudes and risk assessments. A descriptive study in one institution.* (PhD), University of Birmingham, Birmingham.

Colclough, N. D., Lock, R., & Soares, A. (2011). Pre-service teachers' subject knowledge of and attitudes about radioactivity and ionising radiation. *International Journal of Science Education, 33*(3), 423.

Cooper, S., Yeo, S., & Zadnik, M. (2003). Australian students' views on nuclear issues: Does teaching alter prior beliefs? *Physics Education, 38*(2), 123.

Cors, R. (2016). *Informal science learning: an investigation of how novelty and motivation affect interest development at a mobile laboratory.* (PhD), University of Geneva, Geneva.

Cors, R., Robin, N., Mueller, A., & Kunz, P. (2017). Towards a more comprehensive framework for investigating novelty at out-of-school learning places for science and technology learning. *Progress in Science Education, 1*(1).

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98.

Couch, B. A., Hubbard, J. K., & Brassil, C. E. (2018). Multiple-true-false questions reveal the limits of the multiple-choice format for detecting students with incomplete understandings. *BioScience, 68*(6), 455.

Cox, A. J., & Junkin III, W. F. (2002). Enhanced student learning in the introductory physics laboratory. *Physics Education, 37*(1), 37.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671.

Cronbach, L. J. (1941). An experimental comparison of the multiple true-false and multiple multiple-choice tests. *Journal of Educational Psychology, 32*(7), 533.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297.

Crouch, C., Fagen, A. P., Callan, J. P., & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics, 72*(6), 835.

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105.

Deci, E. L., & Ryan, R. M. (1981). Curiosity and Self-Directed Learning: The Role of Motivation in Education.

DeWitt, J., & Storksdieck, M. (2008). A short review of school field trips: Key findings from the past and implications for the future. *Visitor Studies, 11*(2), 181.

Dierking, L. D., Falk, J. H., Rennie, L., Anderson, D., & Ellenbogen, K. (2003). Policy statement of the "informal science education" ad hoc committee. *Journal of Research in Science Teaching, 40*(2), 108.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American Journal of Public Health, 88*(2), 216.

Dillon, J. (2008). A review of the research on practical work in school science. *King's College, London*.

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics-Physics Education Research, 5*(2), 020103.

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics-Physics Education Research, 2*(1), 010105.

DiSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *The Jean Piaget Symposium Series. Constructivism in the Computer Age*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Driver, R. (1981). Pupils' alternative frameworks in science. *European Journal of Science Education, 3*(1), 93.

Duit, R. (2009). Bibliography STCSE: Students' and teachers' conceptions and science education. University of Kiel.

Duit, R., Gropengiesser, H., Kattmann, U., Komorek, M., & Parchmann, I. (2012). The model of educational reconstruction: A framework for improving teaching and learning science. In *Science Education Research and Practice in Europe*: Brill Sense.

Duit, R., & Treagust, D. F. (2012). How can conceptual change contribute to theory and practice in science education? In *Second International Handbook of Science Education*: Springer.

Eckstein, M. K., Guerra-Carrillo, B., Singley, A. T. M., & Bunge, S. A. (2017). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience, 25*, 69.

Edwards, J., & Power, C. (1990). Role of laboratory work in a national junior secondary science project: Australian Science Education Project (ASEP). In E. Hegarty-Hazel (Ed.), *The Student Laboratory and the Science Curriculum*. London: Routledge.

Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden*. Basel: Beltz Verlag.

Eijkelhof, H., Klaassen, C., Lijnse, P., & Scholte, R. (1990). Perceived incidence and importance of lay-ideas on ionizing radiation: Results of a Delphi-study among radiation-experts. *Science Education, 74*(2), 183.

Eijkelhof, H. M. C. (1990). *Radiation and risk in physics education.* (PhD), Rijksuniversiteit Utrecht, Utrecht.

Engeln, K. (2004). *Schülerlabors.* (PhD), Christian-Albrechts-Universität, Kiel.

Estivill-Castro, V. (2002). Why so many clustering algorithms: a position paper. *SIGKDD explorations, 4*(1), 65.

European Centre for the Development of Vocational Training (2009). *European guidelines for validating non-formal and informal learning*: Office for Official Publications of the European Union.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272.

Fadigan, K. A., & Hammrich, P. L. (2004). A longitudinal study of the educational and career trajectories of female participants of an urban informal science education program. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 41*(8), 835.

Falk, J., & Dierking, L. D. (2000). *Learning from museums: visitor experiences and the making of meaning*: AltaMira Press Walnut Creek, CA.

Falk, J., & Storksdieck, M. (2005). Using the contextual model of learning to understand visitor learning from a science center exhibition. *Science Education, 89*(5), 744.

Falk, J. H., & Dierking, L. D. (2016). *The museum experience revisited*: Routledge.

Falk, J. H., Martin, W. W., & Balling, J. D. (1978). The novel field-trip phenomenon: Adjustment to novel settings interferes with task learning. *Journal of Research in Science Teaching, 15*(2), 127.

Fazzini, T., Fidecaro, G., Merrison, A. W., Paul, H., & Tollestrup, A. (1958). Electron decay of the pion. *Physical Review Letters, 1*(7), 247.

Fechner, S. (2009). *Effects of context-oriented learning on student interest and achievement in chemistry education* (Vol. 95): Logos Verlag Berlin GmbH.

Field, A. (2009). *Discovery Statistics Using SPSS: and sex, and drugs and rock 'n'roll*. Los Angeles: CA: Sage.

Förtsch, C., Werner, S., Dorfner, T., Kotzebue, L. v., & Neuhaus, B. J. (2017). Effects of Cognitive Activation in Biology Lessons on Students' Situational Interest and Achievement. *Research in Science Education, 47*(3), 559.

France, S. L., & Batchelder, W. H. (2015). Maximum likelihood item easiness models for test theory without an answer key. *Educational and Psychological Measurement, 75*(1), 57.

Frenzel, A. C., Goetz, T., Lüdtke, O., Pekrun, R., & Sutton, R. E. (2009). Emotional transmission in the classroom: exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology, 101*(3), 705.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change, 114*, 254.

Galbraith, J., Moustaki, I., Bartholomew, D. J., & Steele, F. (2008). *The analysis and interpretation of multivariate data for social scientists (2nd edition)*: Routledge.

Garrett, R., & Roberts, I. (1982). Demonstration versus Small Group Practical Work in Science Education. A critical review of studies since 1900. *Studies in Science Education, 9*(1), 109.

Gläser-Zikuda, M., & Mayring, P. (2003). A qualitative oriented approach to learning emotions at school. In P. Mayring & C. Rhoeneck (Eds.), *Learning Emotions: The Influence of Affective Factors on Classroom Learning*.

Glen, S. (2016). Test-Retest Reliability / Repeatability *Statistics How To. Statistics for the rest of us!* Retrieved from https://www.statisticshowto.datasciencecentral.com/test-retest-reliability/ on 14/03/2020

Glowinski, I. (2007). *Schülerlabore im Themenbereich Molekularbiologie als Interesse fördernde Lernumgebungen.* (PhD), Christian-Albrechts Universität Kiel.

Glowinski, I., & Bayrhuber, H. (2011). Student Labs on a University Campus as a Type of Out-of-School Learning Environment: Assessing the Potential to Promote Students' Interest in Science. *International Journal of Environmental and Science Education, 6*(4), 371.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 1*(4), 223.

Grossnickle, E. M. (2016). Disentangling curiosity: Dimensionality, definitions, and distinctions from interest in educational contexts. *Educational Psychology Review, 28*(1), 23.

Grube, J. W., Morgan, M., & Kearney, K. A. (1989). Using self-generated identification codes to match questionnaires in panel studies of adolescent substance use. *Addictive Behaviors, 14*(2), 159.

Guderian, P. (2007). *Wirksamkeitsanalyse außerschulischer Lernorte.* (PhD), Humboldt-Universität, Berlin.

Guilford, J. P. (1950). *Fundamental statistics in psychology and education*. New York and London: McGraw-Hill Book Company.

Gunstone, R. F., & Champagne, A. B. (1990). Promoting conceptual change in the laboratory. *The Student Laboratory and the Science Curriculum*, 159.

Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. *Education Endowment Foundation*.

Hafele, A. (2012). *Exploring learning difficulties associated with understanding ionizing by radiation.* Paper presented at the National Conference of Undergraduate Research 2012.

Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, 83.

Han, J., Kamber, M., & Tung, A. K. (2001). Spatial clustering methods in data mining. *Geographic Data Mining and Knowledge Discovery*, 188.

Haney, P., & Durlak, J. A. (1998). Changing self-esteem in children and adolescents: A meta-analytical review. *Journal of Clinical Child Psychology, 27*(4), 423.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/ performance measures. *Review of Educational Research, 52*(1), 123.

Hänze, M., & Berger, R. (2007). Cooperative learning, motivational effects, and student characteristics: An experimental study comparing cooperative learning and direct instruction in 12th grade physics classes. *Learning and Instruction, 17*(1), 29.

Harlen, W. (1999). *Effective Teaching of Science. A Review of Research. Using Research Series, 21*: ERIC.

Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments? *Journal of Research in Science Teaching, 37*(7), 655.

Hart, S. G. (2006). *NASA-task load index (NASA-TLX); 20 years later.* Paper presented at the Human factors and ergonomics society annual meeting.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology* (Vol. 52, pp. 139): Elsevier.

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology, 17*(3), 300.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association, 72*(358), 320.

Hattie, J. (2008). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*: Routledge.

Haupt, O. J., Domjahn, J., Martin, U., Skiebe-Corrette, P., Vorst, S., Zehren, W., & Hempelmann, R. (2013). Schülerlabor–Begriffsschärfung und Kategorisierung. *MNU, 66*(6), 324.

Helmke, A. (2002). *Das Projekt MARKUS: Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*: Verlag Empirische Pädagogik.

Hempelmann, R., & Haupt, O. J. (2014). *The German Schülerlabor–Definition and Categories*.

Henriksen, E. K., & Jorde, D. (2001). High school students' understanding of radiation and the environment: Can museums play a role? *Science Education, 85*(2), 189.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141.

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review, 1*(2), 69.

Hidi, S., & Baird, W. (1986). Interestingness-A neglected variable in discourse processing. *Cognitive Science, 10*(2), 179.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111.

Hirth, M. (2019). *Akustische Untersuchungen mit dem Smartphone und Tablet-Computern: Fachliche und didaktische Aspekte.* (PhD), Technische Universität Kaiserslautern, Kaiserslautern.

Hochberg, K. (2016). *iMechanics: Smartphones als Experimentiermittel im Physik-unterricht der Sekundarstufe II Wirkung auf Lernerfolg, Motivation und Neugier in der Mechanik.* (PhD), Technische Universität Kaiserslautern, Kaiserslautern.

Hochberg, K., & Kuhn, J. (2019). What do scientists do? Increasing Awareness of social and networking aspects in everyday activities of scientists. *Progress in Science Education, 2*(1).

Hodson, D. (1988). Experiments in science and science teaching. *Educational Philosophy and Theory, 20*(2), 53.

Hodson, D. (1990). A critical look at practical work in school science. *School Science Review, 71*(256), 33.

Hodson, D. (1993). Re-thinking old ways: Towards a more critical approach to practical work in school science. *Studies in Science Education, 22*(1), 85.

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. *Journal of Curriculum Studies, 28*(2), 115.

Hofer, S. I., Schumacher, R., Rubin, H., & Stern, E. (2018). Enhancing physics learning with cognitively activating instruction: A quasi-experimental classroom intervention study. *Journal of Educational Psychology*.

Hoffmann, L., Häußler, P., & Peters-Haft, S. (1997). *An den Interessen von Mädchen und Jungen orientierter Physikunterricht*. Christian-Albrechts-Universität zu Kiel: IPN. 155.

Hofstein, A., & Kind, P. M. (2012). Learning In and From Science Laboratories. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second International Handbook of Science Education*: Springer.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88*(1), 28.

Hofstein, A., Shore, R., & Kipnis, M. (2004). Providing high school chemistry students with opportunities to develop learning skills in an inquiry-type laboratory: A case study. *International Journal of Science Education, 26*(1), 47.

Hopf, M. (2007). *Problemorientierte Schülerexperimente*: Logos Verlag.

Houseal, A. K., Abd-El-Khalick, F., & Destefano, L. (2014). Impact of a student-teacher-scientist partnership on students' and teachers' content knowledge, attitudes toward science, and pedagogical practices. *Journal of Research in Science Teaching, 51*(1), 84.

Hox, J. (1998). *Multilevel Modeling: When and Why*, Berlin, Heidelberg.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and Applications*: Routledge.

Huwer, J. (2015). *Nachhaltigkeit und Chemie im Schülerlabor: forschendes Experimentieren im Kontext einer naturwissenschaftlich-technischen Umweltbildung.* Universität des Saarlandes, Saarbrücken.

Hynd, C. R., McWhorter, J. Y., Phares, V. L., & Suttles, C. W. (1994). The role of instructional variables in conceptual change in high school physics topics. *Journal of Research in Science Teaching, 31*(9), 933.

Itzek-Greulich, H. (2015). *Einbindung des Lernorts Schülerlabor in den naturwissenschaftlichen Unterricht: Empirische Untersuchung zu kognitiven und motivationalen Wirkungen eines naturwissenschaftlichen Lehr-Lernarrangements.* Eberhard Karls Universität, Tübingen.

Itzek-Greulich, H., Flunger, B., Vollmer, C., Nagengast, B., Rehm, M., & Trautwein, U. (2015). Effects of a science center outreach lab on school students' achievement–Are student lab visits needed when they teach what students can learn at school? *Learning and Instruction, 38*, 43.

Itzek-Greulich, H., Flunger, B., Vollmer, C., Nagengast, B., Rehm, M., & Trautwein, U. (2017). Effectiveness of lab-work learning environments in and out of school: A cluster randomized study. *Contemporary Educational Psychology, 48*, 98.

Jansky, A. (2019). *Die Rolle von Schülervorstellungen zu Wahrscheinlichkeit und Zufall im naturwissenschaftlichen Kontext. Konzeption und Evaluation einer Lehr-Lern-Einheit zur Wahrscheinlichkeitstheorie in der Radioaktivität.* (PhD), Universität Wien.

Johnson, A. (2013). Radiation and atomic literacy for nonscientists. *Science, 342*(6157), 436.

Johnson, A. (2014). Inquiry into Radioactivity (IiR). Retrieved on 14/03/2020 from www.camse.org/radiation

Johnson, A., & Hafele, A. (2010). *Exploring Student Understanding Of Atoms And Radiation With The Atom Builder Simulator.* Paper presented at the AIP Conference 2010.

Johnson, A., & Maidl, R. (2014). *Students Coming to Understand Ionizing Radiation-A Radiation Literacy Challenge.* Paper presented at the National Association for Research in Science Teaching 2014 Annual Conference, Pittsburgh PA.

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. *Journal of Computer Assisted Learning, 7*(2), 75.

Jurik, V., Gröschner, A., & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences, 32*, 132.

Kaczmarek, R., Bednarek, D., & Wong, R. (1987). Misconceptions of medical students about radiological physics. *Health Physics, 52*(1), 106.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963.

Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological testing: principles, applications, and issues*: Nelson Education.

Kashdan, T. B., & Silvia, P. J. (2009). Curiosity and interest: The benefits of thriving on novelty and challenge. *Oxford Handbook of Positive Psychology, 2*, 367.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344): John Wiley & Sons.

Keller, M., Neumann, K., & Fischer, H. E. (2013). Teacher enthusiasm and student learning. *International Guide to Student Achievement*, 247.

Kelley, T., Ebel, R., & Linacre, J. (2002). Item discrimination indices. *Rasch Measurement Transactions, 16*(3), 883.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75.

Klaassen, C., Eijkelhof, H., & Lijnse, P. (1990). Considering an alternative approach to teaching radioactivity. In *Relating macroscopic phenomena to microscopic particles: a central problem in secondary science education* (pp. 304). Utrecht: Cd-B Press.

Klaassen, C. W. J. M. (1995). *A problem-posing approach to teaching the topic of radioactivity.* (PhD), University Utrecht, Utrecht.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik (Ed.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137). Münster: Waxmann.

Kline, T. J. (2005). Classical Test Theory: Assumptions, Equations, Limitations, and Item Analyses. *Psychological testing: a practical approach to design and evaluation*, 107.

Knobloch, S., Patzig, G., Mende, A.-M., & Hastall, M. (2004). Affective news: Effects of discourse structure in narratives on suspense, curiosity, and enjoyment while reading news and novels. *Communication Research, 31*(3), 259.

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12*(4), 383.

Krapp, A., Renninger, A., & Hidi, S. (1992). Interest, learning, and development. In A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development.* Hillsdale: NJ: Erlbaum.

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research, 33*(2), 188.

Kuhn, J. (2010). Authentische Aufgaben im theoretischen Rahmen von Instruktions-und Lehr-Lern-Forschung. *Optimierung von Ankermedien für eine neue Aufgaben-kultur im Physikunterricht. Wiesbaden: Vieweg+ Teubner*.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2013). *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers Results from the COACTIV Project*: Springer Science & Business Media.

Kunter, M., Tsai, Y.-M., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learning and Instruction, 18*(5), 468.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159.

Langford, I. H., & Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 161*(2), 121.

Levene, H. (1961). Robust tests for equality of variances. *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*, 279.

Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology, 74*, 150.

Li, J., & Singh, C. (2017). Developing and validating a conceptual survey to assess introductory physics students' understanding of magnetism. *European Journal of Physics, 38*(2), 025702.

Lijnse, P., Eijkelhof, H., Klaassen, C., & Scholte, R. (1990). Pupils' and mass-media ideas about radioactivity. *International Journal of Science Education, 12*(1), 67.

Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie*. Berlin: Springer.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction, 19*(6), 527.

Litman, J., Hutchins, T., & Russon, R. (2005). Epistemic curiosity, feeling-of-knowing, and exploratory behaviour. *Cognition & Emotion, 19*(4), 559.

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment, 80*(1), 75.

Lord, F. M. (1952). The relation of the reliability of multiple-choice tests to the distribution of item difficulties. *Psychometrika, 17*(2), 181.

Lorenzo-Seva, U., & Ferrando, P. J. (2012). TETRA-COM: A comprehensive SPSS program for estimating the tetrachoric correlation. *Behavior Research Methods, 44*(4), 1191.

Lowry, N., & Johnson, D. W. (1981). Effects of controversy on epistemic curiosity, achievement, and attitudes. *The Journal of Social Psychology, 115*(1), 31.

Lyons, T., & Quinn, F. (2010). *Choosing Science: Understanding the declines in senior high school science enrolments*: National Centre of Science, ICT and Mathematics Education for Rural and Regional Australia.

Maas, C. J., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis, 46*(3), 427.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1*(4), 173.

Mahalanobis, P. C. (1936). *On the generalized distance in statistics.* Paper presented at the National Institute of Science, India.

Maidl, R., & DeKay, N. (2012). *Identifying and resolving problematic student reasoning about ionizing radiation.* Paper presented at the National Conference of Undergraduate Research 2012.

Maiseyenka, V. (2014). *Modellbasiertes Experimentieren im Unterricht: Praxistauglichkeit und Lernwirkungen* (Vol. 166): Logos Verlag.

Mallery, P., & George, D. (2003). *SPSS for Windows step by step: a simple guide and reference*. Boston: Allyn & Bacon.

Maloney, D. P. (1985). Charged poles? *Physics Education, 20*(6), 310.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics, 69*(1), 12.

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*(4), 623.

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*(4), 542.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self‐concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*(2), 397.

Martin, L. M. (2004). An emerging research framework for studying informal learning and schools. *Science Education, 88*(1).
Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology, 100*(2), 291.
McKagan, S., Handley, W., Perkins, K., & Wieman, C. (2009). A research-based curriculum for teaching the photoelectric effect. *American Journal of Physics, 77*(1), 87.
Mehmetoglu, M., & Jakobsen, T. G. (2016). *Applied statistics using Stata: a guide for the social sciences*: Sage.
Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science, 159*(3810), 56.
Meyer, D. K., & Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist, 37*(2), 107.
Millar, R. (1993). *School students' understanding of key ideas about radiation and radioactivity.* (Science Education Research Paper), University of York, York.
Millar, R. (1994). School students' understanding of key ideas about radioactivity and ionizing radiation. *Public Understanding of Science, 3*, 53.
Millar, R., & Gill, J. S. (1996). School students' understanding of processes involving radioactive substances and ionizing radiation. *Physics Education, 31*(1), 27.
Millar, R., Klaassen, K., & Eijkelhof, H. (1990). Teaching about radioactivity and ionising radiation: an alternative approach. *Physics Education, 25*(6), 338.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.
Miller, K., Lasry, N., Chu, K., & Mazur, E. (2013). Role of physics lecture demonstrations in conceptual learning. *Physical Review Special Topics - Physics Education Research, 9*(2), 020113.
Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*(3), 424.
Mokhonko, S. (2016). Nachwuchsförderung im MINT-Bereich.
Mokhonko, S., Nickolaus, R., & Windaus, A. (2014). Förderung von Mädchen in Naturwissenschaften: Schülerlabore und ihre Effekte. *Zeitschrift für Didaktik der Naturwissenschaften, 20*(1), 143.
Molyneux-Hodgson, S., Sutherland, R., & Butterfield, A. (1999). *Is 'Authentic'Appropriate? The Use of Work Contexts in Science Practical Activity.* Paper presented at the International Conference on Practical Work, Copenhagen.
Molz, A. (2016). *Verbindung von Schülerlabor und Schulunterricht-Auswirkungen auf Motivation und Kognition im Fach Physik.* (PhD), Technische Universität Kaiserslautern.
Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105.
Mubeen, S. M., Abbas, Q., & Nisar, N. (2008). Knowledge about ionising and non-ionising radiation among medical students. *Journal of Ayub Medical College, 20*(1), 118.
Muijs, D. (2004). *Doing quantitative research in education with SPSS*: Sage Publications.
Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), *Gender and occupational outcomes: Longitudinal assessments of individual, social, and cultural influences*: American Psychological Association.
Nagy, G., Watt, H. M., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *Journal of Research on Adolescence, 20*(2), 482.
Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects

models revisited and expanded. *Journal of the Royal Society Interface, 14*(134), 20170213.

National Research Council. (2009). *Learning science in informal environments: people, places, and pursuits*. Washington, DC: National Academies Press.

Naylor, F. D. (1981). A state-trait curiosity inventory. *Australian Psychologist, 16*(2), 172.

Neumann, S., & Hopf, M. (2012). Students' Conceptions About 'Radiation': Results from an Explorative Interview Study of 9th Grade Students. *Journal of Science Education and Technology, 21*(6), 826.

Neumann, S., & Hopf, M. (2013). Children's drawings about "radiation"—before and after Fukushima. *Research in Science Education, 43*(4), 1535.

Niaz, M. (1995). Cognitive conflict as a teaching strategy in solving chemistry problems: A dialectic–constructivist perspective. *Journal of Research in Science Teaching, 32*(9), 959.

Nissen, J. M., Talbot, R. M., Thompson, A. N., & Van Dusen, B. (2018). Comparison of normalized gain and Cohen's d for analyzing gains on concept inventories. *Physical Review Physics Education Research, 14*(1), 010115.

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41*(3), 181.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity, 41*(5), 673.

O'Sullivan, J., O'Connor, O. J., O'Regan, K., Clarke, B., Burgoyne, L. N., Ryan, M. F., & Maher, M. M. (2010). An assessment of medical students' awareness of radiation exposures associated with diagnostic imaging investigations. *Insights into Imaging, 1*(2), 86.

OECD (2006). *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006*.

OECD (2007). *PISA 2006: Science Competencies for Tomorrow's World, Volume 1: Analysis*.

OECD (2014). Indicator D1: How much time do students spend in the classroom? In *Education at a Glance 2014: OECD Indicators*: OECD Publishing.

OECD Global Science Forum (2006). *Evolution of student interest in science and technology studies: Policy report.*

Oh, J.-Y. (2011). Using an enhanced conflict map in the classroom (photoelectric effect) based on lakatosian heuristic principle strategies. *International Journal of Science and Mathematics Education, 9*(5), 1135.

Orion, N., & Hofstein, A. (1991). Factors Which Influence Learning Ability during a Scientific Field Trip in a Natural Environment. *Journal of Research in Science Teaching, 31*(10).

Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of Research in Science Teaching, 31*(10), 1097.

Orion, N., Hofstein, A., Tamir, P., & Giddings, G. J. (1997). Development and validation of an instrument for assessing the learning environment of outdoor science activities. *Science Education, 81*(2), 161.

Osborne, J., & Dillon, J. (2007). Research on learning in informal contexts: Advancing the field? Taylor & Francis.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049.

Osborne, J. W., Costello, A. B., & Kellow, J. T. (2008). Best Practices in Exploratory Factor Analysis. In J. Osborne (Ed.), *Best Practices in Quantitative Methods* (pp. 86): Sage Publications.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193.

Parker, J. M., Anderson, C. W., Heidemann, M., Merrill, J., Merritt, B., Richmond, G., & Urban-Lurain, M. (2012). Exploring undergraduates' understanding of photosynthesis using diagnostic question clusters. *CBE—Life Sciences Education, 11*(1), 47.

Parker, P., Nagy, G., Trautwein, U., & Lüdtke, O. (2014). Predicting career aspirations and university majors from academic ability and self-concept: a longitudinal application of the internal–external frame of reference model. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment: A life course perspective* (pp. 224): Cambridge University Press.

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. *Psychological Bulletin, 134*(2), 270.

Pawek, C. (2009). *Schülerlabore als interessefördernde außerschulische Lernumgebungen für Schülerinnen und Schüler aus der Mittel-und Oberstufe.* (PhD), Christian-Albrechts Universität Kiel.

Pestalozzi, J. H. (1826). *Sämmtliche Schriften: Pestalozzi's Schwanengesang. 13* (Vol. 13): Cotta.

Pirker, C. (2009). *Statistical noise or valuable information: The role of extreme cases in marketing research*: Springer Science & Business Media.

Pleijel, H. (1936). *Award ceremony speech of the Nobel Prize in Physics*: Nobel Committee for Physics of the Royal Swedish Academy of Sciences.

Plotz, T. (2017). Students' Conceptions of Radiation and What to Do about Them. *Physics Education, 52*(1), 014004.

Plotz, T., & Hopf, M. (2016). Two Concepts of Radiation: A Case Study Investigating Existing Preconceptions. *European Journal of Science and Mathematics Education, 4*(4), 447.

Podgoršak, E. B. (2014). Compendium to Radiation Physics for Medical Physicists.

Porst, R. (2008). *Fragebogen: Ein Arbeitsbuch*: Springer.

Prather, E. (2005). Students' beliefs about the role of atoms in radioactive decay and half-life. *Journal of Geoscience Education, 53*(4), 345.

Prather, E. E. (2000). *An investigation into what students think and how they learn about ionizing radiation and radioactivity.* (PhD), University of Maine.

Prather, E. E., & Harrington, R. R. (2001). Student understanding of ionizing radiation and radioactivity. *Journal of College Science Teaching, 31*(2), 89.

Priemer, B., Menzl, C., Hagos, F., Musold, W., & Schulz, J. (2018). Das situationale epistemische Interesse an physikalischen Themen von Mädchen und Jungen nach dem Besuch eines Schülerlabors. *Zeitschrift für Didaktik der Naturwissenschaften, 24*(1), 279.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2 Ed. Vol. 1): Sage.

Rego, F., & Peralta, L. (2006). Portuguese students' knowledge of radiation physics. *Physics Education, 41*(3), 259.

Reio Jr, T. G., Petrosko, J. M., Wiswell, A. K., & Thongsukmag, J. (2006). The measurement and conceptualization of curiosity. *The Journal of Genetic Psychology, 167*(2), 117.

Rennie, L., & Howitt, C. (2015). *Evaluation of the Scientists in Schools project 2011-2012*: Report prepared for CSIRO Education.

Rennie, L. J. (1994). Measuring affective outcomes from a visit to a science education centre. *Research in Science Education, 24*(1), 261.

Renninger, K. A. (2007). *Interest and motivation in informal science learning*.

Riesch, W., & Westphal, W. (1975). Modellhafte Schülervorstellungen zur Ausbreitung radioaktiver Strahlung. *Der Physikunterricht, 9*(4), 75.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science education now: A renewed pedagogy for the future of Europe*: European Commission Directorate-General for Research Information and Communication Unit.

Rodenhauser, A. (2018). *Bilinguale biologische Schülerlaborkurse. Konzeption und Durchführung sowie Evaluation der kognitiven und affektiven Wirksamkeit.* (PhD), Universität Wuppertal.

Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: Thirst for knowledge. *Learning and Instruction, 32*, 37.

Rotgans, J. I., & Schmidt, H. G. (2018). How individual interest influences situational interest and how both are related to knowledge acquisition: A microanalytical investigation. *The Journal of Educational Research, 111*(5), 530.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin, 28*(12), 1629.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68.

Ryan, R. M., & Deci, E. L. (2016). Facilitating and Hindering Motivation, Learning, and Well-being in Schools: Research and Observations from Self-Determination Theory. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of Motivation at School* (2nd Edition): Routledge.

Scaife, T. M., & Heckler, A. F. (2010). Student understanding of the direction of the magnetic force on a charged particle. *AmericanJournal of Physics, 78*(8), 869.

Scaife, T. M., & Heckler, A. F. (2011). Interference between electric and magnetic concepts in introductory physics. *Physical Review Special Topics - Physics Education Research, 7*(1), 010104.

Scharfenberg, F.-J. (2005). *Experimenteller Biologieunterricht zu Aspekten der Gentechnik im Lernort Labor: empirische Untersuchung zu Akzeptanz, Wissenserwerb und Interesse.* (PhD), Universität Bayreuth.

Schecker, H., Wilhelm, T., Hopf, M., & Duit, R. (2018). *Schülervorstellungen und Physikunterricht: Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis*: Springer-Verlag.

Schiefele, H. (1978). Lernmotivation und Motivlernen: Grundzüge einer erziehungswissenschaftlichen Motivationslehre. *München: Ehrenwirth*.

Schiefele, U. (1991). Interest, learning, and motivation. *Educational psychologist, 26*(3-4), 299.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development*: Hillsdale, NJ : Erlbaum.

Schmitt, F. F., & Lahroodi, R. (2008). The epistemic value of curiosity. *Educational Theory, 58*(2), 125.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*(1), 23.

Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (2n Edition ed., pp. 34): Routledge.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461.

Simon, F. (2019). *Der Einfluss von Betreuung und Betreuenden auf die Wirksamkeit von Schuelerlaborbesuchen.* (PhD), TU Dresden

Sjøberg, S., & Schreiner, C. (2010). *The ROSE project: An overview and key findings*. Retrieved on 14/03/2020 from https://roseproject.no/?p=63

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles, 50*(3-4), 241.

Snijders, T. A., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research, 22*(3), 342.

StataCorp. (2019). *Stata 16 Base Reference Manual*. College Station: TX: Stata Press.

Streller, M. (2015). The educational effects of pre and post-work in out-of-school laboratories (PhD). Technical University Dresden.

Student. (1908). The probable error of a mean. *Biometrika*, 1.

Stylinski, C., Storksdieck, M., Canzoneri, N., Klein, E., & Johnson, A. (2018). Impacts of a comprehensive public engagement training and support program on scientists' outreach attitudes and practices. *International Journal of Science Education, Part B, 8*(4), 340.

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching, 49*(4), 515.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295.

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*: Thomson/Brooks/Cole.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.): Pearson.

Taber, K. S. (2013). Revisiting the chemistry triplet: drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice, 14*(2), 156.

Taber, K. S. (2017). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 1.

Tamir, P. (1989). Some issues related to the use of justifications to multiple-choice answers. *Journal of Biological Education, 23*(4), 285.

Taslidere, E. (2016). Development and use of a three-tier diagnostic test to assess high school students' misconceptions about the photoelectric effect. *Research in Science & Technological Education, 34*(2), 164.

Tekin, B. B., & Nakiboglu, C. (2006). Identifying students' misconceptions about nuclear chemistry. A study of Turkish high school students. *Journal of Chemical Education, 83*(11), 1712.

Todt, E., & Schreiber, S. (1998). Development of interests. In Hoffmann, L., Krapp, A., Renninger, K.A., & Baumert, J. (Eds.) *Interest und learning. Proceedings of the Seeon Conference on Interest and Gender.* Kiel: IPN.

Trumper, R. (1997). Applying conceptual conflict strategies in the learning of the energy concept. *Research in Science & Technological Education, 15*(1), 5.

Trumper, R. (2003). The physics laboratory–a historical overview and future perspectives. *Science & Education, 12*(7), 645.

Tytler, R., Symington, D., & Clark, J. C. (2017). Community-school collaborations in science: Towards improved outcomes through better understanding of boundary issues. *International Journal of Science and Mathematics Education, 15*(4), 643.

Ukoumunne, O. C., Gulliford, M. C., & Chinn, S. (2002). A Note on the Use of the Variance Inflation Factor for Determining Sample Size in Cluster Randomized Trials. *Journal of the Royal Statistical Society. Series D (The Statistician), 51*(4), 479.

Urban, D., & Mayerl, J. (2006). *Regressionsanalyse: Theorie, Technik und Anwendung*: VS Verlag für Sozialwissenschaften.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111.

Van Dusen, B., & Nissen, J. (2019). Modernizing use of regression models in physics education research: A review of hierarchical linear modeling. *Physical Review Physics Education Research, 15*(2), 020108.

Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS One, 8*(7).

Von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science, 6*(6), 574.

Waternaux, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika, 63*(3), 639.

Wavo, E. n.-Y.-T. (2004). Honesty, cooperation and curiosity and achievement of some schools on Nanjing (China). *IFE PsychologIA : An International Journal, 12*(2).

Wellington, J. (1990). Formal and informal learning in science: The role of the interactive science centres. *Physics Education, 25*(5), 247.

Wentzel, K. R., & Miele, D. B. (2016). *Handbook of Motivation at School* (2nd Edition): Routledge.

Weßnigk, S. (2012). *Kooperatives Arbeiten an industrienahen außerschulischen Lern-orten.* (PhD), Christian-Albrechts Universität Kiel.

Wherry, R. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *The annals of mathematical statistics, 2*(4), 440.

White, R., & Gunstone, R. (1992). *Probing understanding*. London: Routledge.

Wiener, G. (2017). *Elementary particle physics in early physics education.* (PhD), Universität Wien.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental review, 12*(3), 265.

Wild, E., & Möller, J. (2009). *Pädagogische Psychologie*: Springer.

Wild, K. P. (2000). *Der Einfluss von Unterrichtsmethoden und motivationalen Orientierungen auf das kognitive Engagement im Berufsschulunterricht.* Paper presented at the Workshop: Ergebnisse fachdidaktischer und psychologischer Lehr-Lern-Forschung, Pädagogischen Hochschule Ludwigsburg.

Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594.

Woithe, J. (2016). *CLOUD CHAMBER. S'Cool LAB - Do-it-yourself manual.* Retrieved on 14/03/2020 from https://scoollab.web.cern.ch/cloud-chamber

Woods-Townsend, K., Christodoulou, A., Rietdijk, W., Byrne, J., Griffiths, J. B., & Grace, M. M. (2016). Meet the scientist: The value of short interactions between scientists and students. *International Journal of Science Education, Part B, 6*(1), 89.

Woolnough, B. E., & Allsop, T. (1985). *Practical work in science*: Cambridge University Press.

Yunker, M. L. (2010). *A Systemic Examination of the Introduction of an Outdoor Learning-Based Science Curriculum to Students, their Teacher, and the School Principal.* (PhD), University of Michigan, Michigan.

Zehren, W. (2009). *Forschendes Experimentieren im Schülerlabor.* (PhD), Universität des Saarlandes, Saarbrücken.

Zeldin, A. L., Britner, S. L., & Pajares, F. (2008). A comparative study of the self‐efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 45*(9), 1036.

Zwick, M. M., & Renn, O. (2000). *Die Attraktivität von technischen und ingenieurwissenschaftlichen Fächern bei der Studien-und Berufswahl junger Frauen und Männer*: Akademie für Technikfolgenabschätzung in Baden-Württemberg.

12 Acknowledgments

The development of S'Cool LAB, all its education content, its accompanying research, and this dissertation would not have been possible without the help of numerous supportive advisors, colleagues and friends. I look back at an exceptionally educational and challenging chapter of my professional life, which was composed of gratifying learning opportunities both about physics education research and about myself, enlightening interactions with students, teachers, tutors and fellow researchers from around the world, dreadful, yet joyful encounters with CERN's fire fighters, but most of all by the collaborative spirit at CERN in which I was fortunate to take part.

From the very beginning, I was expertly supported by my university advisors, Prof. Dr. Jochen Kuhn and Prof. Dr. Andreas Müller, who not only helped me in developing a deeper understanding of out-of-school learning opportunities and their challenges and opportunities for physics education research, but also guided me throughout my doctoral research project with their extensive knowledge and experience. In particular, Prof. Dr. Jochen Kuhn took on the challenge of advising my project remotely and always took time out of his busy schedule to help me with clear and concise feedback. I am particularly indebted to Prof. Dr. Andreas Müller. Throughout months-long struggles to get S'Cool LAB up and running, he remained immensely patient and supportive, but he also continuously demanded progress, which forced me to balance the demands of S'Cool LAB's management and this dissertation project. He was always available to provide thorough, clear, and helpful advice, but he also made sure I kept a positive attitude towards my research project and my data. Thank you, Andreas!

I also want to thank my CERN supervisor, Dr. Sascha Schmeling, for his continuous support, trust and the incredible opportunities that allowed me to grow, both personally and professionally. I am very grateful for these opportunities, and I am looking forward to our future collaboration.

I am also very grateful for the generous donation by the Lotterie Romande, who made the S'Cool LAB project possible, and for the tremendous work by Dr. Jeff Wiener, who expertly managed the design and construction of S'Cool LAB and presented me with a beautiful work environment. Consequently, I am grateful for every single emergency intervention of the CERN fire brigade, who protected S'Cool LAB from flooding, collapsing rainwater pipes and teenagers, dead frogs and falling trees. Furthermore, I want to thank all the students who devoted their time to participate in my studies.

Over the years, S'Cool LAB grew into a diverse team of fantastic colleagues, who all contributed to its success in one way or the other. Some of them stayed only for a month to develop a new hands-on workshop, others accompanied S'Cool LAB over years and majorly shaped its existence. I would like to thank, in alphabetical order, Alexandra, Alex, Chris, Christoph, Daniela, Dimitrios, Eva, everyone else I forgot to list here, Fabian, Hashim, Kitti, Lachlan, Lea, Lukas, Magnus, Mihaly, Nathalia, Nick, Oliver, Sara, Sarah, Stavie, Susanne, Tiago, Vincent and Zuzana. Furthermore, I would like to express my sincere gratitude to all CERN scientists, who took time out of their busy and important work schedules to facilitate workshops in S'Cool LAB. Indeed, all the amazing S'Cool LAB tutors passionately and patiently introduced more than 30 000 students to the fascinating world of particle physics. There is a reason why you were awarded by one of the highest regression coefficients within the S'Cool LAB project, thank you very much!

My time at CERN was greatly influenced by my encounter with Dr. Alexandra Jansky, who spent hundreds of hours with me in S'Cool LAB developing and testing new activities, entertaining numerous and diverse groups of students, and laughing about things that went differently as expected. It is impossible to find a better colleague and friend and I look forward to our continued collaboration!

I want to thank my entire family − especially my parents, Gabriele and Frank Woithe − for their life-long support and backing. Last, I want to express my special gratitude to Jeff for his invaluable personal and scientific companionship over the past years and for his immense patience, support and encouragement. I am truly grateful to have you in my life!

13 Curriculum vitae

passionate mathematics and physics teacher, creative problem solver and education content designer, project manager, science education researcher

PERSONAL DETAILS

Nationality	German
Address	Rue Dr-Alfred-Vincent 12, CH-1201 Genève
Contact details	+41754116539, julia.woithe@cern.ch, linkedin.com/in/julia-woithe
Date of birth	03/12/1987 in Löbau, Germany

PROFESSIONAL EXPERIENCE

09/2017 – today — **Senior Research Fellow**, International Relations Department, Teacher and Student Programmes Section, CERN, Geneva, Switzerland

- Management of CERN's hands-on particle physics learning laboratory "S'Cool LAB" for high-school students from all around the world
  - Development and empiric evaluation of inquiry-based activities
  - Facilitation of workshops for high-school students and teachers
  - Handling workshop requests for 8000 participants per year
  - Management of S'Cool LAB team and 50 volunteers
  - Budget and personnel planning
  - Organisation of residential high-school student programmes
  - Supervision of PhD & master students
  - Public representation of S'Cool LAB at conferences & VIP visits

03/2014 – 08/2017 — **PhD Student**, CERN, Geneva, Switzerland & TU Kaiserslautern, Germany

- Wolfgang Gentner Scholarship by German Federal Ministry of Education and Research (BMBF)
- Development, management, and quantitative evaluation of S'Cool LAB

08/2013 – 02/2014 — **CERN Project Associate,** "Netzwerk Teilchenwelt", TU Dresden, Germany

- Organisation of residential programmes for high-school students & teachers
- Development of a particle physics curriculum and teaching resources

05/2008 – 07/2013 — **Student Research Assistant**, OncoRay, TU Dresden, Germany

- Assistance with conducting and documenting experiments with laser accelerated particle beams (Laser-Radiooncology)
- Data analysis, contributions to publications, literature research

09/2008 – 12/2012 — **Mathematics Tutor**, "Studienkreis Radebeul", Germany

- Tutoring students between grades 3 and 12 in small groups

LANGUAGES

German	Native proficiency (C2)
English	Full professional proficiency (C2)
French	Limited working proficiency (B1)

EDUCATION

03/2014 – 04/2020	**PhD Student, TU Kaiserslautern, based at CERN, Geneva** PhD thesis: Designing, Measuring and Modelling the Impact of the Hands-on Particle Physics Learning Laboratory S'Cool LAB at CERN.
04/2012 – 07/2013	**Master of Education, TU Dresden (Mathematics & Physics, ECTS grade A)** Master's thesis: Analysis of the interpretation of ATLAS data by students in the context of the International Masterclasses in Particle Physics 2013.
10/2009 – 03/2012	**Bachelor of Education, TU Dresden (Mathematics & Physics)** Bachelor's thesis: Kerr-Microscopy: Development of a hands-on workshop for the high-school student learning laboratory DeltaX (HZDR).
10/2006 – 09/2009	**Physics studies, TU Dresden**
07/2006	**A levels (Abitur)** Sächsisches Landesgymnasium St. Afra Meißen, Germany

SKILLS & INTEREST

Digital tools	SPSS, MS Office, GeoGebra, MATLAB, Plickers
Certificates	Google Certified Educator Level 1, Swiss Radiation Protection Expert B/C
3D printing	Designing and printing educational setups, e.g., experimental equipment
Escape games	Designing and facilitating educational (physics) escape games
Aspiring astronaut	Top30 candidate of "Die Astronautin" (first female German astronaut initiative)
Crafting	Head of CERN's crafting society (knitting, felting, crochet, ...)

14 Publications in the framework of this PhD project

Conference proceedings

- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020) Hands-on & Minds-on Teilchenphysikexperimente im CERN Schülerlabor S'Cool LAB. In: Passon, O., Zügge, T., Grebe-Ellis, J. (Eds.). Kohärenz im Unterricht der Elementarteilchenphysik. Tagungsband des Symposions zur Didaktik der Teilchenphysik in Wuppertal 2018. Springer Verlag
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2018). Konzeptuelles Verständnis im Schülerlabor. In: Maurer, C. (Eds.). Qualitätsvoller Chemie- und Physikunterricht- normative und empirische Dimensionen. 38. Tagungsband der Gesellschaft für Didaktik der Chemie und Physik
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2016). Konzepttest für Teilchenphysik im Forschungsumfeld Schülerlabor. In: Maurer, C. (Eds.). Authentizität und Lernen - das Fach in der Fachdidaktik. 35. Tagungsband der Gesellschaft für Didaktik der Chemie und Physik

Other conference contributions

- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Educational Outcomes of a Hands-on Particle Physics Learning Laboratory. Conference of DiNat - Swiss Science Education Association, 24/01/2020, Genf, talk
- Woithe, J., Kuhn, J., Müller, A., & Schmeling, S. (2019) SchülerInnenvorstellungen zu ionisierenden Teilchen und Radioaktivität. GDCP conference, 09-12/09/2019, Vienna, poster
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2019). Affective and Cognitive Effects of Hands-on Particle Physics Workshops – An International Study. NARST Conference 2019, Baltimore, 31/03-03/04/2019, talk
- GDCP conference, 18-21/09/2017, Regensburg, talk and proceedings
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2016). Hands-on Particle Physics Experiments for High-school Students at S'Cool LAB / CERN. GIREP seminar, 30/08-03/09/2016, Krakow, talk
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2016). Students' Conceptions and Hands-on Learning of Particle Physics in S'Cool LAB. Conference of DINAT - Swiss Science Education Association, 29/01/2016, Fribourg, poster
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2015). CERN S'Cool LAB – Ein internationales Teilchenphysik-Schülerlabor. DPG spring meeting, Wuppertal, 09-13/03/2015, poster

Planned papers in peer-reviewed journals

- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Multi-level modelling of affective outcomes of an out-of-school learning lab. To be submitted to: Journal of Research in Science Teaching
- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Factor-cluster segmentation of students' conceptions about radiation and radioactivity. To be submitted to: Physical Review Physics Education Research
- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Conceptual learning in the framework of an out-of-school learning lab. To be submitted to: Progress in Science Education

All publications & conference contributions: cern.ch/julia.woithe

15 Lebenslauf

Enthusiastische Mathematik- und Physiklehrerin,
Kreative Problemlöserin und Designerin von Lernmaterialien,
Projektmanagerin, Physikdidaktik-Forscherin

ANGABEN ZUR PERSON

Nationalität	Deutsch
Adresse	Rue Dr-Alfred-Vincent 12, CH-1201 Genf, Schweiz
Kontakt	+41754116539, julia.woithe@cern.ch, linkedin.com/in/julia-woithe
Geburtsdatum	03/12/1987 in Löbau, Deutschland

BERUFSERFAHRUNG

09/2017 – Heute

Senior Research Fellow, International Relations Department, Teacher and Student Programmes, CERN, Genf, Schweiz

- Management von CERNs Teilchenphysik-Schülerlabor “S’Cool LAB” für Schülerinnen und Schüler aus aller Welt
  - Entwicklung und empirische Evaluation von Teilchenphysik-Experimenten und Lernaktivitäten
  - Durchführung von Teilchenphysik-Workshops Schülerinnen und Schüler sowie deren Lehrpersonen
  - Management von Buchungsanfragen (8000 Teilnehmende pro Jahr)
  - Management des S'Cool LAB Team and von 50 Freiwilligen
  - Budget- und Personalplanung
  - Organisation von Camps
  - Betreuung von Doktorierenden & Masterstudierenden
  - Öffentliche Repräsentation von S'Cool LAB bei Konferenzen und VIP Besuchen

03/2014 – 08/2017

Doktorandin, CERN, Genf, Schweiz & TU Kaiserslautern, Deutschland

- Wolfgang Gentner Stipendium des Bundesministeriums für Bildung und Forschung (BMBF)
- Entwicklung, Management und quantitative Evaluation von S'Cool LAB

08/2013 – 02/2014

CERN Project Associate, "Netzwerk Teilchenwelt", TU Dresden, Deutschland

- Organisation von CERN Programmen für Jugendliche und Lehrpersonen
- Entwicklung von Unterrichtsmaterialien zur Teilchenphysik in Zusammenarbeit mit Lehrkräften

05/2008 – 07/2013

Studentische Hilfskraft, OncoRay, TU Dresden, Deutschland

- Mitarbeit an der Durchführung und Auswertung von medizinphysikalischen Experimenten und an wissenschaftlichen Publikationen

09/2008 – 12/2012

Mathematik Nachhilfelehrerin, “Studienkreis Radebeul”, Deutschland

- Hausaufgabenbetreuung in Kleingruppen und individuelle Nachhilfe für Jugendliche der Jahrgangsstufen 3 bis 12

SPRACHEN

Deutsch	Muttersprache (C2)
Englisch	Verhandlungssicher (C1/C2)
Französisch	Sehr gute Kenntnisse (B1)

SCHUL- UND BERUFSBILDUNG

03/2014 – 04/2020	**Doktorandin, TU Kaiserslautern, stationiert am CERN, Genf** Dissertation: Designing, Measuring and Modelling the Impact of the Hands-on Particle Physics Learning Laboratory S'Cool LAB at CERN.
04/2012 – 07/2013	**Master of Education, TU Dresden (Mathematik & Physik, ECTS grade A)** Höheres Lehramt an Gymnasien für Mathematik und Physik Masterarbeit: Analyse der Auswertung von ATLAS-Daten durch Jugendliche im Rahmen der Internationalen Masterclasses der Teilchenphysik 2013
10/2009 – 03/2012	**Bachelor of Education, TU Dresden (Mathematik & Physik)** Bachelorarbeit: Kerr-Mikroskopie – Entwicklung einer Experimentiersequenz im Schülerlabor DeltaX
10/2006 – 09/2009	**Physikstudium, TU Dresden**
07/2006	**Abitur** Sächsisches Landesgymnasium St. Afra Meißen, Deutschland

FÄHIGKEITEN & INTERESSEN

Digitalkompetenz	SPSS, MS Office, GeoGebra, MATLAB, Plickers
Zertifikate	Google Certified Educator Level 1, Swiss Radiation Protection Expert B/C
3D-Druck	3D-Design und 3D-Druck von Experimentiermaterialien
Escape Games	Gestalten und Durchführen von pädagogischen (Physik-) Escape Games
Raumfahrt	Top30 Kandidatin bei “Die erste deutsche Astronautin“
Basteln	Leiterin des CERN Bastelclubs (Stricken, Häkeln, Filzen, ...)

16 Publikationen im Rahmen dieser Doktorarbeit

Beiträge in Tagungsbänden

- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020) Hands-on & Minds-on Teilchenphysikexperimente im CERN Schülerlabor S'Cool LAB. In: Passon, O., Zügge, T., Grebe-Ellis, J. (Eds.). Kohärenz im Unterricht der Elementarteilchenphysik. Tagungsband des Symposions zur Didaktik der Teilchenphysik in Wuppertal 2018. Springer Verlag
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2018). Konzeptuelles Verständnis im Schülerlabor. In: Maurer, C. (Eds.). Qualitätsvoller Chemie- und Physikunterricht- normative und empirische Dimensionen. 38. Tagungsband der Gesellschaft für Didaktik der Chemie und Physik
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2016). Konzepttest für Teilchenphysik im Forschungsumfeld Schülerlabor. In: Maurer, C. (Eds.). Authentizität und Lernen - das Fach in der Fachdidaktik. 35. Tagungsband der Gesellschaft für Didaktik der Chemie und Physik

Andere Konferenzbeiträge

- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Educational Outcomes of a Hands-on Particle Physics Learning Laboratory. Conference of DiNat – Verband Fachdidaktik Naturwissenschaften Schweiz, 24/01/2020, Genf, Vortrag
- Woithe, J., Kuhn, J., Müller, A., & Schmeling, S. (2019) SchülerInnenvorstellungen zu ionisierenden Teilchen und Radioaktivität. GDCP Tagung, 09-12/09/2019, Wien, Poster
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2019). Affective and Cognitive Effects of Hands-on Particle Physics Workshops – An International Study. NARST Conference 2019, Baltimore, 31/03-03/04/2019, Vortrag
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2016). Hands-on Particle Physics Experiments for High-school Students at S'Cool LAB / CERN. GIREP seminar, 30/08-03/09/2016, Krakau, Vortrag
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2016). Students' Conceptions and Hands-on Learning of Particle Physics in S'Cool LAB. Conference of DINAT - Swiss Science Education Association, 29/01/2016, Fribourg, Poster
- Woithe, J., Schmeling, S., Kuhn, J., & Müller, A. (2015). CERN S'Cool LAB – Ein internationales Teilchenphysik-Schülerlabor. DPG Frühjahrstagung, Wuppertal, 09-13/03/2015, Poster

Geplante referierte Veröffentlichungen (peer-reviewed)

- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Multi-level modelling of affective outcomes of an out-of-school learning lab. Wird eingereicht bei: Journal of Research in Science Teaching
- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Factor-cluster segmentation of students' conceptions about radiation and radioactivity. Wird eingereicht bei: Physical Review Physics Education Research
- Woithe, J., Kuhn, J., Müller, A., Schmeling, S. (2020). Conceptual learning in the framework of an out-of-school learning lab. Wird eingereicht bei: Progress in Science Education

Alle Publikationen & Konferenzbeiträge: cern.ch/julia.woithe