Received: 9 April 2021 | Revised: 27 November 2021 | Accepted: 8 January 2022

DOI: 10.1002/tea.21748

RESEARCH ARTICLE

JRST | WILEY

Motivational outcomes of the science outreach lab S'Cool LAB at CERN: A multilevel analysis

Julia Woithe[1] | **Andreas Müller**[2] | **Sascha Schmeling**[1] | **Jochen Kuhn**[3]

[1]CERN, Geneva, Switzerland

[2]Faculty of Science/Physics Section and Institute of Teacher Education, University of Geneva, Geneva, Switzerland

[3]Department of Physics, Physics Education Research Group, University of Kaiserslautern, Kaiserslautern, Germany

Correspondence
Julia Woithe, CERN, CH-1211 Geneva, Switzerland.
Email: julia.woithe@cern.ch

Funding information
German Federal Ministry of Education and Research, Wolfgang-Gentner Scholarship

Abstract

Previous studies highlight the positive effects of science outreach labs, in particular on students' motivational variables. However, out-of-school learning is generally associated with high novelty and specific setting characteristics that can impact learning and development. Indeed, previous studies call for further research on students' perception of the learning settings to ensure the best possible use of science outreach labs. This study aims to take this call up by analyzing motivational outcomes (situational interest and self-concept) together with an unprecedented number of carefully chosen student and setting factors supposed to contribute to students' experience at science outreach labs. This study involved 509 high-school students from 13 countries who took part in a half-day hands-on session at the particle physics outreach lab S'Cool LAB at CERN and a single group longitudinal pre- and post-test research design. The results confirm that this intervention led to very high situational interest and self-concept, even for a student sample that showed higher-than-average dispositional interest and self-concept beforehand. Moreover, the initial motivational gender gap was closed after the intervention. To take the nested data into account, multilevel models were employed to study the predictive power of a set of student factors as well as students' perception of setting factors. Here, even after controlling for student factors such as their dispositional interest, support by the learning environment and educators was a crucial setting factor and was associated with especially high situational interest. Furthermore, students' cognitive preparedness and cognitive load were vital with respect to their situational self-concept. Overall, regression models account for almost 60% of the variance of both motivational outcomes. We conclude that a systematic measurement of student and setting factors together with a multilevel approach provides highly valuable information about science outreach labs and how to optimize their effectiveness.

KEYWORDS

interest, multilevel modeling, out-of-school science learning, science outreach lab, self-concept

1 | INTRODUCTION

Although many students find science and technology important and hold a mainly positive attitude toward science, technology, engineering, and mathematics (STEM) subjects, it seems that formal science education has failed to harness students' positive attitudes (OECD, 2006; Sjøberg & Schreiner, 2010). In most of the OECD countries, the number of students opting for science and technology has declined over recent decades, and women still remain underrepresented in STEM (OECD Global Science Forum, 2006; UNESCO, 2017). In particular, students' declining interest in school science, their low science self-concept, their stereotypic images of science and scientists, and their resulting low interest in science and technology careers are problematic (Osborne et al., 2003; UNESCO, 2017).

This trend is particularly worrying since today's students, both female and male, will require scientific literacy and scientific reasoning skills to face the various challenges of the 21st century, for example, the revolution of job markets due to machine learning and robotics. Moreover, science and engineering occupations have low probability of computerization as they require higher-order complex thinking skills and creative intelligence and are thus among the safest careers (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 decrease this worrying decline of students' interest in science and science careers. For example, Rocard et al. (2007) highlight the importance of fostering students' self-concept in science, especially for female students. Others recommend more hands-on activities and inquiry-based learning, as well as more field trips and stronger links between formal and informal science education (Boiko et al., 2019; Rocard et al., 2007). Lyons and Quinn (2010) suggest bringing students in contact with scientists, for example, under the framework of out-of-school learning opportunities.

Indeed, out-of-school science learning opportunities[1] have demonstrated a strong development over the past decades. Today, numerous science museums, science centers, and various research institutions offer a plethora of out-of-school science learning opportunities, which are increasingly being recognized worldwide as an essential part of STEM education (Corrigan et al., 2018; National Research Council, 2009; Singh, 2009; Werquin, 2010). Friedman (2010) provides an overview of the evolution of science centers and museums toward the interactive, hands-on approach they adopt nowadays, which is considered as particularly promising for audiences of high school learners (Braund & Reiss, 2006; Stocklmayer et al., 2010). The "science outreach lab" (Goldschmidt & Bogner, 2016; Thomas, 2012) is a unique out-of-school science learning opportunity that focuses on active hands-on experimentation activities, and is usually offered by science departments of universities, research institutions, and science centers (Exploratorium, 2020; Fermilab, 2020; Glowinski & Bayrhuber, 2011; Heureka, 2020). This study employs a multilevel approach to investigate the effectiveness and mechanisms of action of the science outreach lab "S'Cool LAB" (CERN, 2020) with respect to the motivational variables interest and self-concept. The remainder of the paper is structured as follows. We first review the roles of interest and self-concept in out-of-school science learning. Then, we describe the science outreach lab S'Cool LAB at CERN, before introducing factors that can hinder learning and development at science outreach labs. We next describe the methods of the study, present and discuss its results, and finish with conclusions and outlook.

2 | RESEARCH BACKGROUND

2.1 | The motivational variables interest and self-concept

In the following, we describe interest and self-concept as motivational variables using Deci and Ryan's empirical self-determination theory (SDT) as a framework to link interest and self-concept to learning motivation. In particular, interest is described as an important source of intrinsic motivation and also as a regulatory process for intrinsically motivated behavior (O'Keefe et al., 2017; Ryan & Deci, 2000). Moreover, the need for competence satisfaction in self-determination theory is very closely related to the operationalization of self-concept (Marsh et al., 2017).

On the one hand, interest refers to a psychological state aroused by interactions between persons and their environment (e.g., objects, topics, contents, ideas). This "situational interest" is characterized by focused attention and an affective reaction, such as enjoyment (Hidi & Baird, 1986; Hidi & Renninger, 2006; Krapp, 2002; Schiefele, 1978). On the other hand, interest is characterized by a relatively long-lived predisposition to engage with certain ideas or learning content, and is often referred to as "individual," "personal," or "dispositional interest."

Although situational and dispositional interests are different constructs, it is assumed that a high situational interest triggered by specific characteristics of a learning environment can transform into more stable forms of interest under the right conditions (Krapp, 2002). In this sense, out-of-school learning activities affect both stable and malleable components of interest. For example, stable interest in a topic correlates positively with conceptual understanding of physics texts, independent of students' prior knowledge (Andre & Windschitl, 2003), and with conceptual change (Mason et al., 2008). However, even students that show the more short-lived situational interest demonstrate a higher persistence and deeper processing of information, which in turn are associated with improved comprehension and learning (Ainley, Hidi, & Berndorff, 2002; Ainley, Hillman, & Hidi, 2002; Schraw & Lehman, 2001). Moreover, interest is an important aspect of course and career choices (Hidi, 2006; Krapp, 2002; Schraw & Lehman, 2001).

However, previous research indicates that students' interest in science (Osborne et al., 2003) and other academic topics (Todt & Schreiber, 1998; Wigfield & Eccles, 1992) decreases during adolescence. Moreover, this effect seems to be stronger for girls than for boys (Hoffmann et al., 1997; Marginson et al., 2013). Out-of-school learning opportunities can be used to foster students' interest by including activities triggering high situational interest (Glowinski & Bayrhuber, 2011; Itzek-Greulich & Vollmer, 2017), independently of students' previous knowledge, dispositional interests, and gender. Ultimately, repeated arousal of situational interest can increase students' dispositional interests (Rotgans & Schmidt, 2017). For example, Goff et al. (2020) report a high positive correlation between undergraduates' interest in science and math and their prior experience of out-of-school science learning opportunities.

The self-concept of ability is a mental representation of oneself that includes a collection of cognitive concepts about oneself formed through experience and feedback from the environment (Gutman & Schoon, 2013; Valentine et al., 2004). Thus, domain-specific self-concept (e.g., academic self-concept) refers to perceptions of oneself in a specific area. Marsh (1990) provides evidence for a hierarchical multi-faceted academic self-concept and recommends using subject-specific scales when measuring a specific component of academic self-concept.

The construct of academic self-concept shares many similarities with academic self-efficacy, and self-efficacy may act as a precursor to the development of self-concept (Bong & Skaalvik, 2003; Marsh et al., 2017). While self-concept describes relatively stable perceptions oriented at past experiences, self-efficacy, a concept originally proposed by Bandura (1977), embodies more malleable conceptions that are oriented at future context-specific tasks (Bandura, 1977; Bong & Skaalvik, 2003). However, despite the slight difference in their theoretical conceptualization, it is methodologically challenging to measure academic self-concept and self-efficacy beliefs separately (Valentine et al., 2004). Moreover, Valentine et al. (2004) report no differences between self-efficacy measures and the domain-specific academic self-concept. In our study, we focus on the self-concept in physics. Nevertheless, in the following, we briefly review research on both academic self-concept and self-efficacy that we consider relevant to out-of-school learning keeping in mind that self-efficacy may indeed act as precursor to self-concept.

Research within formal education contexts has demonstrated reciprocal effects between academic self-concept and achievement, and recommended fostering both simultaneously (Marsh & O'Mara, 2008; Valentine et al., 2004). Additionally, academic self-concept is an important predictor of students' high-school course choices and career aspirations (Nagy et al., 2008; Parker et al., 2014). Pajares and Miller (1994) describe self-concept as mediating the influence of gender and prior experience on mathematics performance. However, students' self-concept declines with age and is found to be lower for women than for men (science: OECD (2007); mathematics: Nagy et al. (2010)). Also, women and men seem to differ in how they report on sources of their science self-efficacy beliefs with women reporting, for example, social persuasion more often than men (Usher et al., 2019; Zeldin et al., 2008). Moreover, the role of the sources of science self-efficacy depends on the educational environment (Dorfman & Fortus, 2019).

In principle, out-of-school learning activities can foster students' science self-concept and self-efficacy beliefs through challenging but achievable tasks, feelings of competence, and social persuasion. Although Goff et al. (2020) report a high positive correlation between undergraduates' competence beliefs in science and math and their prior experience of out-of-school science learning opportunities, our understanding of the development of self-concept in out-of-school learning settings is still limited (Goff et al., 2020; Schunk & DiBenedetto, 2016). Here, research on out-of-school learning settings can contribute by investigating, for example, factors leading to self-concept development, to ultimately help improve our understanding of motivational variables such as self-concept in general.

2.2 | The science outreach lab "S'Cool LAB"

This article focuses on the effects of the science outreach lab "S'Cool LAB" at CERN as one example of out-of-school learning. Goldschmidt and Bogner (2016) describe science outreach labs as science laboratories at universities or research institutions that are dedicated to education and outreach and can be visited by school classes. Depending on the host institution, science outreach labs focus, for example, on biology, chemistry, physics, general science, and technology-related activities. Furthermore, these labs, which are often funded by outreach budgets of large research institutions, provide authentic learning environments, are typically related to modern science topics, and offer research equipment that is usually not available in schools (Glowinski & Bayrhuber, 2011; Goldschmidt & Bogner, 2016; Itzek-Greulich et al., 2015; Thomas, 2012). This combination of characteristics puts science outreach labs at a unique position within the out-of-school learning sector, although individual components of science outreach labs, such as hands-on activities or the involvement of scientists, are also offered in other learning settings.

The science outreach lab "S'Cool LAB" is located at the particle physics research laboratory CERN in Geneva, Switzerland, and focuses on the modern physics topic particle physics. S'Cool LAB offers opportunities to engage in hands-on activities using authentic particle physics research equipment under the guidance of volunteers from CERN's scientific community.

Which characteristics make science outreach labs such as "S'Cool LAB" promising out-of-school science learning settings? Habig et al. (2020) describe important design principles of successful out-of-school science learning programs. For example, interviewees highlighted the importance of interaction with STEM professionals, which is in line with the results by Fadigan and Hammrich (2004) who report that scientists in out-of-school settings can greatly influence (female) students' educational and career decisions. Moreover, Habig et al. (2020) highlight the value of providing authentic learning experiences that allow learners to become practitioners of science, for example, through laboratory investigations. Indeed, students prefer practical work to theoretical work (Swarat et al., 2012) and girls especially seem to benefit from hands-on experiences (Burkam et al., 1997). Even if this may be limited to situational interest and does not necessarily imply a longer-term personal interest in science (Abrahams, 2009), this is certainly a positive factor, especially for learners who have not had many opportunities to experience situations that promote interest in science. Furthermore, Krapp and Prenzel (2011) discuss how repeated experience of situational interest can evolve into personal interest.

Obviously, the affective and cognitive effectiveness of lab work depend on its quality. Effective lab work, for example, concentrates on few but important experimental tasks (Hodson, 1993; Hofstein et al., 2005; Hofstein & Mamlok-Naaman, 2007), aims for the right amount of cognitive load by balancing openness and instructional guidance (Kirschner et al., 2006; Müller & Brown, 2022), and enables learners to reflect on their prior conceptions (Cinici & Demir, 2013; Gunstone & Champagne, 1990; Miller et al., 2013). A review on the topic can be found in Lunetta et al. (2007), and a summary of meta-analytic results about experimental tasks in Müller and Brown (2022).

Science outreach labs are popular not only among students and teachers but also among education and outreach coordinators of universities and research institutions, which is reflected in the very high number of labs, almost 400 in Germany alone, as listed in Lernort Labor (2020); the international development is described in the Introduction.

Indeed, previous studies confirm positive effects of science outreach labs on affective variables such as interest and self-concept (Itzek-Greulich et al., 2017; Pawek, 2009; Priemer et al., 2018; Rodenhauser, 2018). However, the type of the examined interventions as well as the size of their educational outcomes varies. Of course, any outcomes of an educational intervention strongly depend on students' dispositional interest and self-concept related to the topic of the intervention. Students' interest in biology topics tends to be higher than in chemistry or physics topics, both of which inspire similar interest in students (OECD, 2007). Consequently, comparisons between different science outreach labs must be interpreted with caution. Nevertheless, in the following, we report selected results of studies on science outreach labs with different foci. For simplified comparison and interpretation, Cohen et al. (1999) suggest the linear transformation of raw scores to percent of maximum possible (POMP) scores, which are defined as $\frac{S_o - S_{\min}}{S_o - S_{\max}} \times 100\%$, where S_o is the observed score for a single case, while $S_{\min}$ and $S_{\max}$ are the minimum and maximum possible score for a given scale, respectively. When assessing interest using rating scales, students are usually asked to indicate the level of agreement or disagreement with interest-related statements. For these types of bidirectional rating scales, a POMP score of 0% would indicate strong disinterest, 50% neutral opinion, and 100% a very high interest. In this POMP metric, scores of situational interest after taking part in science outreach labs show a wide range. For example, Itzek-Greulich et al. (2017) report a situational interest score that corresponds to a POMP value of 56% indicating that students' interest was barely triggered by the science outreach lab. However, in their study, students followed cookbook-like instructions to conduct easy chemistry experiments involving starch. Consequently, the authors recommend "a teaching style that cognitively activates the heterogeneity of all students involved" for future studies (Itzek-Greulich et al., 2017). A much higher POMP score of 70% was reported by Pawek (2009) indicating a considerable situational interest of students who took part in science and technology experiments at one of the well-equipped labs of the German Aerospace Center (DLR). Additionally, in a large study with more than 11,000 school students from grades 1 to 13 taking part in science outreach labs that focused on physics topics, Priemer et al. (2018) found an average POMP score of 59% for the epistemic component of situational interest after students took part in physics experiments. In their study, the level of situational interest also declined with age: students in the last 3 years of high-school (grades 11–13, aged 17–19 years) demonstrated the lowest situational interest with a POMP score of 44% indicating rather slight disinterest after taking part in activities at a science outreach lab.

Furthermore, Rodenhauser (2018) observed that different biological activities of the same science outreach lab lead to different educational outcomes. This variation might mirror a varying quality of the learning experiences or their associated educational potential. Previous studies have also shown that the added value of science outreach labs can be limited when comparing their cognitive and affective outcomes with 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 science outreach labs. In summary, not all interventions at science outreach labs seem to make full use of their unique potential and educators at science outreach labs should not solely rely on their special setting but also implement findings from education research, for example, on hands-on learning.

2.3 | Novelty and other factors influencing outcomes of science outreach labs

Falk et al. (1978) first suggested that the novelty of out-of-school learning settings influences students' behavior and cognition. Orion and Hofstein (1994) describe two main factors influencing the educational effectiveness of field trips: quality of the learning opportunity and the multi-dimensional "novelty space," differentiating different dimensions of novelty, such as cognitive novelty or novelty related to the location of an out-of-school learning offer. Moreover, Palmer (2009) identified novelty as a main source of interest in science lessons. Itzek-Greulich and Vollmer (2017) argue that the novelty aspect and the authentic real-world scientific facilities make science outreach labs especially potent to trigger favorable motivational outcomes. However, too much novelty is not entirely beneficial, because it might also distract and overwhelm students (Itzek-Greulich et al., 2015).

We next list different factors that have previously been shown to influence out-of-school learning, many of which can also be interpreted from a novelty perspective.

2.3.1 | Research evidence for factors influencing out-of-school learning

Factors influencing the educational outcomes of out-of-school science learning opportunities include characteristics of learners (student factors) such as prior knowledge, experiences, and interests, as well as the characteristics of the out-of-school science learning opportunity and its perception (setting factors) such as orientations or perceived support (Falk & Storksdieck, 2005; Orion & Hofstein, 1994). Falk and Storksdieck (2005) argue that no single factor alone influences learning outcomes, rather it is the result of a combination of several interacting factors. Moreover, they suggest that future research "should collect specific data for a very homogeneous subset of visitors, increase the sample size, improve our tools for discriminating visitors, or all of the above" to reach a better understanding of individual effects of these factors. In the following, we provide a rationale for the student and setting factors included in the present study, specifically as predictors of motivational outcomes, and link previous research on out-of-school learning offers with related evidence from educational research.[2]

2.3.2 | Student factors

Interest and self-concept in physics

First, it is important to measure the initial level of students' motivational variables because the positive development of motivation is a core objective of out-of-school learning, and because it is a potential predictor of this development. Highly motivated students might benefit more from a visit to a science outreach lab than less motivated ones in the sense of the motivational

"Matthew" (or cumulative advantage) effect (Walberg & Tsai, 1983). For science (and mathematics) education, Durik et al. (2015) review accumulating evidence showing that the effectiveness of instructional interventions in promoting interest depends on students' levels of individual interest and self-concept of ability. Bong et al. (2015) found prior individual interest to be the strongest determinant of self-efficacy development. Moreover, students' self-concept develops through experience and feedback from the environment (Gutman & Schoon, 2013; Valentine et al., 2004). We thus included pre-intervention measures of *dispositional interest* and *self-concept of ability* in the analysis

Curiosity

Second, epistemic curiosity is an important variable at the affective–cognitive intersection (von Stumm et al., 2011), and belongs to the very rationale of many science outreach offers as seen by research, scientists, and providers alike, from early classical papers (Bettelheim, 1980; Oppenheimer, 1975) to a broad literature from the last decades (Falk & Dierking, 2000; National Research Council, 2009; Skydsgaard et al., 2016; Stocklmayer et al., 2010). Moreover, epistemic curiosity is strongly associated with the construct of interest (Alexander & Grossnickle, 2016). Therefore, students' curiosity state was considered as important variable for the analysis.

Prior experiences

Third, students' prior experiences were considered as important factors as they might affect how students perceive a new experience at a science outreach lab, and thus influence their motivational outcomes. Although hands-on experimentation plays a central role in science education (Hofstein & Lunetta, 2004), the number of experiments in physics classrooms varies across countries (Börlin, 2012). Moreover, students with more hands-on experience in physics might feel more comfortable and less overwhelmed by experimental activities at a science outreach lab. Therefore, students' *experience with hands-on experimentation* was included in the analysis. Additionally, *language skills in English*, the working language of S'Cool LABLAB at CERN, were also taken into account. Language skills affect how well students can interact with English worksheets including experiment instructions, and English-speaking tutors. Low English skills might add extraneous cognitive load and hinder learning and development. Furthermore, out-of-school learning places might be overwhelming, simply because some students are not used to these settings. Indeed, students' *experience of previous out-of-school science learning* has been shown to predict educational outcomes (Cors, 2016) and was thus included in the analysis. In the framework of this study's intervention, students were confronted with particle physics, a topic that is not always part of physics curricula. Previous *experience of particle physics as a learning topic* might have reduced students' perceived cognitive novelty, and it was thus taken into account in the analysis.

Finally, *gender* and *age* were considered as control variables. Indeed, compared to male students, female students tend to show lower interest in physics (Häussler & Hoffmann, 2002; Hoffmann et al., 1997) and have lower self-concept in certain subjects (Nagy et al., 2010). In large scale studies, gender differences for interest and self-concept (in favor of boys/men) were found to be small, but non-zero [interest: $d = 0.4$, for the 5000 grade 10 students as part of the UPMAP study in England (Mujtaba & Reiss, 2013); self-concept: $d = 0.27$, OECD average for science in general (OECD, 2007)]. Moreover, both interest and self-concept are known to decline with age (Todt & Schreiber, 1998; Wigfield & Eccles, 1992). Indeed, previous research identified age as an important predictor of motivational outcomes at a science outreach lab (Rodenhauser & Preisfeld, 2018).

Setting factors
Students' perception of their science outreach lab experience was also considered as a so-called "setting factor" due to its potential effect on educational outcomes (Cors et al., 2017; Molz, 2016). Perceived support through organizational features and tutors has been identified as an important predictor of the motivational effects of out-of-school science offers by previous studies (Orion & Hofstein, 1994; Pawek, 2009). Moreover, students' feeling of preparedness and ease of orientation were considered as a "setting orientation" factor. This factor reflects students' perception of novelty related to the unusual setting of an out-of-school science learning opportunity and was linked to the educational effectiveness of field trips by Orion and Hofstein (1994). Due to our working memory's limited capacity, a high amount of extraneous cognitive load can hinder learning and development (Sweller, 1994). The workshops in the science outreach lab presented in this study involved high levels of complexity caused, for example, by the equipment used or the physics principles. Therefore, students' perception of problems occurring during the experiments and cognitive load related to the experiments were considered in the analysis.

2.4 | Motivation for the present study and research questions

Science outreach labs might be relatively expensive outreach tools, but they have an immense potential to influence students' motivation through unique learning experiences. However, previous studies on science outreach labs suffer from severe limitations: first, with only a small number of well-controlled studies, and the absence of replication or meta-analytic studies (Chi et al., 2015; Hausamann, 2012), there is a need to better understand the essential impact factors of out-of-school learning offers, both at student and setting level. Second, there are limitations at the level of individual studies concerning their design, instruments, lack of control and predictor variables, and underreporting (e.g., not providing effect sizes or sufficient data to compute them), as stated repeatedly by researchers in the field (Chi et al., 2015; Hausamann, 2012; Itzek-Greulich & Vollmer, 2017). Broad reviews by the US National Research Council summarize that is it necessary to further improve the quality of evidence on learning science in informal environments (National Research Council, 2009) and that currently available research findings are "not yet robust enough to determine which programs work best for whom and under what circumstances" (National Research Council, 2015).

In particular, the impact of science outreach opportunities depends on the quality of the intervention, as well as on several influence factors. The mixed results of previous research show a gap regarding the motivational influence factors and their possible interaction.

A recent research project by Itzek-Greulich and co-authors responded impressively to these exhortations to improve methodological quality by using, for example, well-described motivation measures, a large sample, and multilevel analysis (Itzek-Greulich et al., 2017; Itzek-Greulich & Vollmer, 2017). Itzek-Greulich et al. (2017) is the study on motivational outcomes of science outreach labs with the most advanced methodology. However, the actual motivational impact the authors report is limited: while they find a difference in situational interest and perceived competence compared to a control group (without any laboratory work), no, or small and inconsistent differences were found between a science outreach lab and lab-work sessions at school. We suspect that their study did not make full use of the potential of science outreach labs, for example, by offering only simple hands-on learning activities that can also be carried out in a school setting (Itzek-Greulich et al., 2015, 2017). Moreover, while the significance level of many comparisons is assessed, no adjustment for multiple comparisons (Hochberg & Tamhane, 1987) is carried out. Finally, the set of predictor variables in the linear multilevel model by Itzek-Greulich and Vollmer (2017) accounts only for a small amount of modeled variance (19% for interest, 31% for self-competence).

Thus, in line with Durik et al. (2015), that "educators should be aware that motivational enhancements may not work well for all learners" and that evidence about the impact of relevant influence factors is necessary in order to adjust the offer accordingly, we see a research gap (i) for a well-controlled study where evidence for substantial motivational impact of a science outreach lab actually can be provided, and (ii) for an improved and extended set of predictor variables with a larger explanatory power. The present study sets out to fill this gap in two steps.

The first step is to demonstrate the motivational effectiveness of a newly developed science outreach lab on modern science (in particular, particle physics). Here, the specific topic of particle physics allowed first-hand experiences that would otherwise not be accessible in usual physics classrooms at school—one of the main advantages of science outreach labs (Braund & Reiss, 2006). Moreover, the intervention was developed based on recommendations on hands-on work and was facilitated by specifically trained scientists. It was thus expected that the intervention would lead to comparatively high outcomes as it makes full use of the educational potential of out-of-school science learning. This study thus aims to overcome the limitations of Itzek-Greulich and Vollmer (2017) who describe motivational outcomes of an intervention based on easy cookbook type chemistry experiments using only standard lab equipment.

In a second step, this study systematically examines the relationship between the motivational outcomes interest and self-concept of this particle physics outreach lab and a wide range of student and setting factors included based on previous empirical evidence. We aim to systematically derive a synthesis of impact factors that explain a large proportion of the variance of the educational effectiveness of an out-of-school learning intervention. This set of factors will then allow us to study and understand out-of-school learning settings in general in more detail.

Additionally, many of the above-mentioned studies on science outreach labs took place in Germany or Switzerland, and it is not clear to what degree their findings can be generalized for students from other countries and educational systems. In the present work, data for students from 13 different countries were included.

Specifically, this study sought to answer the following research questions:

RQ 1: *To which extent can half-day hands-on sessions at a science outreach lab trigger students' interest and self-concept in physics?*
RQ 2: *Which student and setting factors predict students' motivational outcomes?*

3 | MATERIALS AND METHODS

3.1 | Description of study setting and design

To answer our research questions while taking the specific characteristics of the learning setting into account, we implemented a single group, pre- and post-test research design. Students filled out online questionnaires approximately 30 days before (pre-test) and 10 days after (post-test) taking part in an intervention at a science outreach lab. A self-generated student ID allowed us to match pre- and post-tests. The motivational outcomes interest and self-concept were assessed as dependent variables, student and setting factors as independent variables, and age and gender as control variables. To study the clustered data structure, we also recorded information about students' visit group, their teacher, and the country of their school. Before the intervention at the science outreach lab, students took part in a 2.5-h guided tour of research facilities at CERN.

The intervention itself consisted of a 4.5-h hands-on session with three workshops of 90 min each and took place in CERN's dedicated particle physics outreach lab S'Cool LAB. In particular, students studied the behavior of electron beams in magnetic fields, assembled their own particle detector, a so-called cloud chamber, before observing tracks of ionizing particles from natural radiation, and studied the detection of x-ray photons (Figure 1).

The learning activities were developed using an iterative design process prior to this study by taking into account students' conceptions of the underlying physics concepts as well as recommendations on hands-on work. In particular, student worksheets served as process worksheets reducing cognitive load while structuring the problem-solving process (Van Merrienboer & Sweller, 2005). Moreover, students followed predict-observe-explain cycles (White & Gunstone, 1992) as these can promote correct experimental observations, which in return foster conceptual learning (Miller et al., 2013). Here, experimental tasks were carefully aligned with research about the respective students' conceptions so that observations would cause cognitive conflict in students holding certain misconceptions. A detailed description of the development of the intervention is provided in Woithe (2020). During the intervention, volunteering tutors from CERN's scientific community guided the students and helped them interpret their observations and discuss their findings. Each tutor supported a group of maximum 12 students, and 47% of the tutors were female. Figure 2 shows an excerpt of the student worksheets used in one of the three 90-min workshops, which included working with pixel detectors at an x-ray machine.

3.2 | Sample

Between October 2016 and June 2017, 781 students filled out the online pre-test and took part in the intervention, whereas 534 students filled out the online post-test; 509 students who filled out both the pre- and post-test were included in the analysis. Age of the students ranged from 16 to 19 years ($M = 17.0$ years, $SD = 0.9$ year), and 36% of the students were female. The students were distributed over 28 groups from 13 mostly European countries (Austria, Czech Republic,

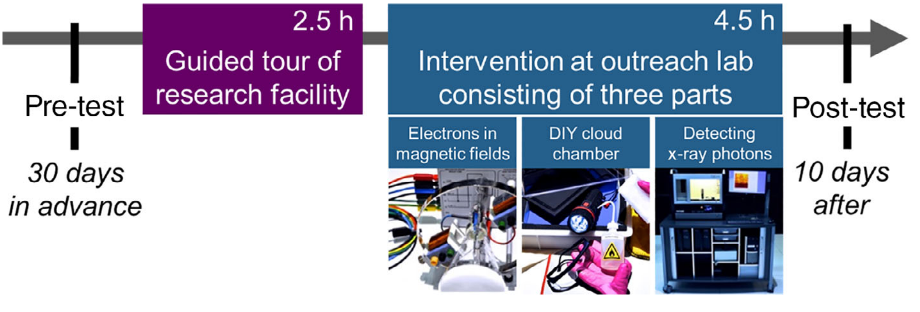
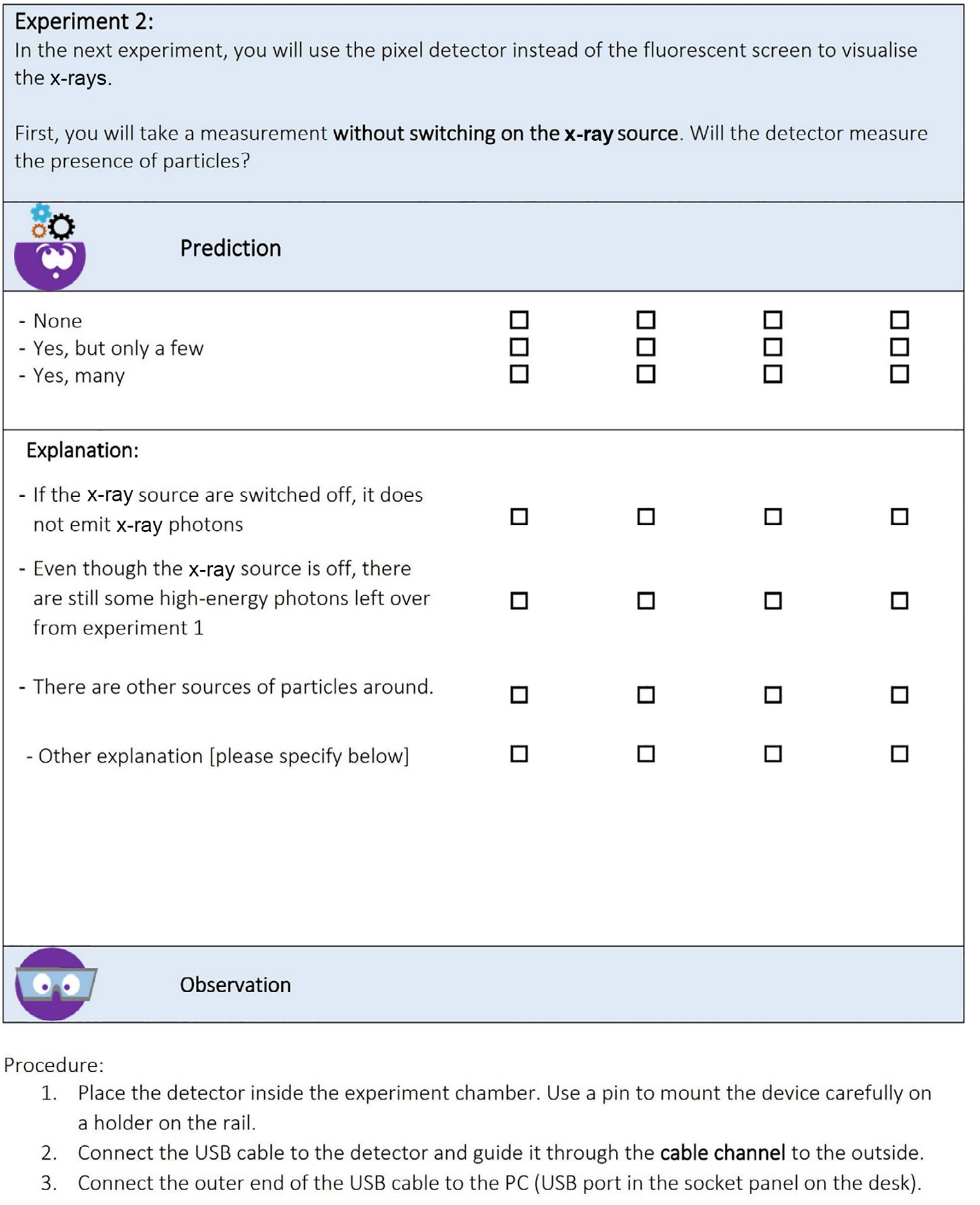


FIGURE 1 Overview of study setting and design



Experiment 2:
In the next experiment, you will use the pixel detector instead of the fluorescent screen to visualise the x-rays.

First, you will take a measurement **without switching on the x-ray source**. Will the detector measure the presence of particles?

Prediction

- None	☐	☐	☐	☐
- Yes, but only a few	☐	☐	☐	☐
- Yes, many	☐	☐	☐	☐

Explanation:

- If the x-ray source are switched off, it does not emit x-ray photons	☐	☐	☐	☐
- Even though the x-ray source is off, there are still some high-energy photons left over from experiment 1	☐	☐	☐	☐
- There are other sources of particles around.	☐	☐	☐	☐
- Other explanation [please specify below]	☐	☐	☐	☐

Observation

Procedure:
1. Place the detector inside the experiment chamber. Use a pin to mount the device carefully on a holder on the rail.
2. Connect the USB cable to the detector and guide it through the **cable channel** to the outside.
3. Connect the outer end of the USB cable to the PC (USB port in the socket panel on the desk).

FIGURE 2 Excerpt of student worksheet of workshop “detecting x-ray photons”

France, Germany, Italy, Norway, the Netherlands, Poland, Portugal, Singapore, Spain, Switzerland, and the UK). Many of the participants followed advanced-level physics courses. Indeed, only 9% of the students reported having less than 120 min of physics lessons per week, 58% of the students had 120–240 min of physics lessons per week, and 34% even reported more than 240 min on physics per week. Moreover, physics was reported as their favorite school subject by more than 40% of the students. More details about the sample are provided in Woithe (2020).

3.3 | Measures

All test instruments were based on well-established scales whose dimensionality and reliability were confirmed by previous studies. However, some items were translated from German to English, adapted to the specific hands-on lab context, or shortened to reduce the test load on the students. Therefore, the properties of the scales were re-assessed in a pilot study, which confirmed their suitability. In particular, principal axis factoring was used to re-confirm the dimensionality of the different scales. Table 1 provides an overview of the psychometric characteristics of the employed scales including their internal consistency if applicable. If not mentioned otherwise, students rated their agreement with statements on a six-level scale from “disagree completely” (1) to “agree completely” (6). Eight of the 10 scales showed a good to very good internal consistency, with Cronbach's $\alpha \geq 0.8$; the remaining two scales showed an adequate internal consistency, with Cronbach's $\alpha \geq 0.7$ (Doran, 1980; Evers et al., 2013). Details about this process and the full list of all item sets are described in Woithe (2020). We next describe the structure and origin of the employed scales in more detail and provide example items.

3.3.1 | Motivational outcomes and dispositional pre-intervention measures

Interest and self-concept in physics

The constructs interest and self-concept in physics were measured based on a well-validated motivation test instrument adapted originally from Hoffmann et al. (1997) and used in previous studies by Hochberg et al. (2018) and Kuhn and Müller (2014). For this study, only items of the interest and self-concept subscales were adapted and translated into English. Factor analysis confirmed the clustering of items within these two subscales. While the items in the pre-test focused on students' dispositional interest in physics (e.g., *I enjoy solving physics problems.*) and dispositional physics self-concept (e.g., *I find it easy to solve tasks in physics classes.*), the items in the post-test assessing the dependent variables focused on students' situational interest (e.g.,

TABLE 1 Overview of scales assessing predictor and motivational outcome variables

Scales (time of measurement)	M	SD	M%	#	α_C	95% CI
Dispositional interest in physics (pre)	4.29	0.94	66%	7	0.87	[0.85, 0.89]
Dispositional physics self-concept (pre)	4.32	1.00	66%	4	0.88	[0.86, 0.90]
Curiosity state particle physics (pre)	4.88	0.95	78%	3	0.88	[0.86, 0.90]
Situational interest in physics (post)	4.84	0.84	77%	7	0.87	[0.85, 0.89]
Situational physics self-concept (post)	4.70	0.77	74%	4	0.83	[0.80, 0.85]
Tinkering support (post)	5.29	0.59	86%	6	0.81	[0.79, 0.84]
Setting orientation (post)	4.62	0.75	72%	5	0.70	[0.66, 0.74]
Lack of problems (post)	4.43	1.09	69%	3	0.78	[0.75, 0.81]
Cognitive load experiments (post)	2.38	0.89	28%	5	0.85	[0.83, 0.87]
Perceived tutor support (post)	5.37	0.68	87%	6	0.90	[0.88, 0.91]

Abbreviations: M, scale mean value; SD, standard deviation; M%, mean value in percent of maximum possible score; #, number of items; α_C, Cronbach's alpha; 95% CI, 95% confidence interval.

During my time in S'Cool LAB, I enjoyed solving physics problems.) and situational self-concept related to the activities in the science outreach lab (e.g., *I found it easy to solve tasks in S'Cool LAB.*). All items were carefully parallelized in order to allow a comparability in the sense of using students' dispositional interest and self-concept as a baseline to interpret the size of their situational interest and self-concept that was triggered by the activities in S'Cool LAB.

A terminological comment is in order here: We have defined interest in an operational way allowing for comparison of its general, long-term perception regarding physics and its perception in the given situation at the science outreach lab. For lack of a better term, we use "dispositional interest" for the former and "situational interest" for the latter. While there is a broader discussion about the distinction and interaction of the two in the literature (Murphy & Alexander, 2000; Rotgans & Schmidt, 2018), this goes beyond the purpose and operationalization of the present paper.

3.3.2 | Additional student factors

To assess students' *curiosity*, items based on Litman and Spielberger (2003) and Naylor (1981), already adapted to and tested in science outreach labs by previous studies in Germany (Hirth, 2019; Hochberg, 2016; Molz, 2016), were used in their original version in English or translated from German into English and adapted to the specific context. The pre-test included three 6-level items to measure students' curiosity toward particle physics (e.g., *"I find it fascinating to spend time on particle physics."*).

To assess students' *prior experience with experiments*, their answers to the question *Experiments: How often do the following activities occur in your physics class?* were analyzed. Students were asked separately about the activities "Individual work" and "Group work" and were provided with seven answer options, from "never" to "every week." Students' answer choices were first transformed to a value indicating the number of hands-on activities per week before adding the values for the activities "Individual work" and "Group work." In total, students reported an average of 0.6 hands-on activities per week. Details about the coding and transformation process as well as the original data can be found in Woithe (2020). The variable was z-transformed before performing regression analysis.

English skills were assessed through the English grade reported by students in the pre-test. Due to the plethora of different grading systems in the respective countries and to take the advantage of native speakers into account, this variable was dichotomized. All native speakers were coded as 1, because it was assumed that English as the working language in the science outreach lab would not offer any additional challenge for them. All students with an English grade of 80% or higher were also coded as 1, because it was assumed that they would be 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).

Similar to their hands-on experience, students rated their *out-of-school science learning experience* answering the question: *How often do you approximately visit the following places or use the following offers?* A sum of students' scores for visits to "museum," "zoo, aquarium, or botanical garden," "science center or science outreach lab," and "universities or research laboratories" was calculated, leading to an average of eight out-of-school science learning opportunities per year. Again, this variable was z-transformed prior to further analysis.

Students were also asked about their *previous experience with particle physics* (*How many hours did you spend on particle physics so far?*) in the pre-test. They were provided with seven answer options from "zero" to "more than 20 hours." Students answered two separate items related to experience "during physics classes" and "outside physics classes." On average, students spent 5.5 h and 3.3 h on particle physics during and outside of their classes, respectively. The sum of their particle physics experience was z-transformed prior to further analysis.

3.3.3 | Setting factors

Molz (2016) developed an instrument to assess students' perception of setting factors in the context of a science outreach lab based on existing items from the Science Outdoor Learning Environment Inventory (SOLEI) instrument (Orion et al., 1997; Yunker, 2010). For this study, suitable items were translated from German into English if not already available in English and adapted to the specific context. Here, the post-test included 14 six-level items. Factor analysis confirmed three sub-scales, namely tinkering support due to organization (e.g., *The instructions were helpful for doing the experiments*.), lack of problems (e.g., *It took a long time until I was familiar with the equipment and was able to start the experiments*.—inversely coded), and setting orientation due to preparation (e.g., *I knew what to expect in the lab*.).

An additional cognitive load scale by Hirth (2019) was adapted for the specific context. In particular, five items referred to the cognitive load of the experiments in general with items focusing, for example, on the perceived difficulty (*The experiments were difficult*.) as well as 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*.).

Finally, six 6-level items based on Pawek (2009) and the SOLEI instrument (Orion et al., 1997; Yunker, 2010) were adapted 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 [...] 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*.).

3.4 | Statistical analyses

3.4.1 | Data preparation and univariate screening

The preparation and initial screening of the data with respect to normal distribution, outliers, and missing data followed the recommendations by Tabachnick and Fidell (2013). Composite and single interval-scaled variables were transformed into z-scores, and their distributions were compared to a normal distribution and scanned for univariate extreme values with respect to the sample mean using the critical z-score $|z| > 3.29$. Due to mandatory fields in the online questionnaires, there were no missing data.

3.4.2 | POMP and effect sizes of group comparisons

As recommended by Cohen et al. (1999), scores of psychometric scales were transformed linearly to POMP scores for simplified interpretation and comparison. We conducted independent-samples *t*-tests to analyze gender differences, as these tests are quite robust against violations of normality assumptions if the sample size is big enough and the two groups similarly large (Muijs, 2004). To quantify statistically significant differences, we calculated the effect size, Cohen's *d*, including the 95% confidence interval as measure of statistical error (Cohen, 1988), for example, $d = 0.5\,[0.4; 0.6]$. We also compared effect sizes to the Hattie "hinge point": Hattie (2008) reports an average effect size of $d = 0.40$, which he uses as a comparison value for the impact of educational interventions.[3]

Similarly, we conducted dependent-samples *t*-tests to compare students' answers between the pre- and post-test to answer Research Question 1 (*To which extent can half-day hands-on sessions at a science outreach lab trigger students' interest and self-concept in physics?*). Corresponding repeated-measures effect sizes were calculated as Cohen's d_{RM}, which uses a specific standard deviation that corrects for the correlation *r* between pre- and post-test values (Morris & DeShon, 2002). Consequently, we report the correlation *r* between pre- and post-test values together with the *t*-test results. In these repeated-measures *t*-tests, we use students' dispositional motivational variables as a baseline to interpret the size of their situational motivational outcomes.

3.4.3 | Linear regression analysis

To answer Research Question 2 (*Which student and setting factors predict students' motivational outcomes?*), linear regression analysis was performed to model the relationships between independent variables (students and setting factors) and the respective dependent variables interest and self-concept. Regression models were calculated on *z*-transformed variables. Consequently, the resulting regression coefficients β_i are already standardized.

First, single-level models were calculated. Here, several assumptions were tested beforehand based on the recommendations by Tabachnick and Fidell (2013) to determine the suitability of data for multiple linear regressions. For example, we excluded multivariate outliers using the Mahalanobis distance (Mahalanobis, 1936) and inspected the residual scatterplot to analyze normality, linearity, and homoscedasticity of the regression residuals and to identify and remove cases that were poorly fit by the regression model.

Due to the high number of independent variables, special care was applied to account for overlapping variance between different predictor variables and to prevent overfitting. In particular, sequential regressions were conducted in which predictors were entered one by one into the model with decreasing predictive power or based on their assumed time of causation (Raudenbush & Bryk, 2002). Consequently, chronologically "older" or more stable variables, such as gender, age, or dispositions entered the model earlier than "newer" and less stable variables, such as personality states or subjective perceptions of the learning setting. Moreover, an adjusted R^2 is reported as model fit parameter because it penalizes a high number of predictors in a model (Wherry, 1931).

3.4.4 | False discovery rate and overfitting

The significance of predictors in a linear model is usually gauged based on the size of the standardized regression coefficients, the corresponding *p*-value, and a certain alpha level such as 5%. However, handling a larger number of independent variables with an unadjusted alpha level can inflate the false discovery rate (Hochberg & Tamhane, 1987). Moreover, one might overfit the model in this way as some predictors can enter the model only "by chance." To limit overfitting in the linear regression models, the Bonferroni correction was applied, which reduces the desirable *p* values by dividing the targeted α-level by the number *m* of conducted significance tests: $\alpha_m = \alpha/m$ (Hox et al., 2017). This correction was also applied when calculating multiple comparisons on the same data set, for example, in the framework of pre–post comparisons.

3.4.5 | Multilevel modeling

Multilevel modeling has several advantages when analyzing clustered data such as data from school contexts as it avoids distorted significance tests through underestimated *p* values (Hox et al., 2017). Indeed, whereas single-level models assume randomly selected individuals, multilevel models assume that groups have been randomly selected from a population of groups. In this study, students were considered level-1 units and student groups level-2 units. The sample of 509 students was distributed over 28 groups, with each group comprising an average of 19 students. Student-level predictors were included as fixed effects in the model and a random intercept term was added to account for group-level differences. The restricted maximum likelihood (REML) estimation was used because it takes the number of fixed effects that need to be estimated into account (Harville, 1977). The assumptions for multilevel linear models are similar to ordinary linear models. Thus, all multilevel models were computed based on the reduced samples in which multivariate outliers had been removed based on the results of the single-level models.

First, "empty" random intercept models without any predictors were computed to estimate the variance components. This allowed to estimate how much of the variance of a certain educational outcome can be accounted for at the student level σ_e^2 (i.e., due to differences between individual students) and at the group level σ_u^2 (i.e., due to differences between groups of students). We then calculated the variance partition coefficient (VPC), also called intra-class coefficient, which gives the proportion of the total variance in the dependent variable that is accounted for by the group level (Goldstein et al., 2002).

$$\mathrm{VPC} = \sigma_u^2 / \left(\sigma_e^2 + \sigma_u^2\right)$$

The VPC can also be used to verify whether multilevel modeling is actually necessary for a given data set. As a rule of the thumb, Mehmetoglu and Jakobsen (2016) recommend multilevel modeling if the VPC exceeds 5%. Similarly, the variance inflation factor (VIF), also called design effect, estimates the effect of variance partitioning when considering clustering of data with an average cluster size *c* (Ukoumunne et al., 2002).

$$\text{VIF} = 1 + (c - 1) \times \text{VPC}$$

The VIF describes by how much the effective 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 and significance levels as stated above.

As a figure of merit of the model fit, we followed the recommendation by Huang (2018) who suggests reporting the "analogue R^2," that is, the reduction in total variance from the null to the full model.

$$\text{Analogue}\, R^2 = 1 - \frac{\sigma^2_{e_{\text{full}}} + \sigma^2_{u_{\text{full}}}}{\sigma^2_{e_0} + \sigma^2_{u_0}}$$

In the following, the size of analogue R^2 is interpreted in the same way as the model fit parameter "adjusted R^2" for single-level regression model, with, for example, $R^2 > 0.5$ indicating good model fit according to Muijs (2004).

4 | RESULTS

4.1 | General motivational outcomes interest and self-concept

First, we present the results related to Research Question 1 (*To which extent can half-day hands-on sessions at a science outreach lab trigger students' interest and self-concept in physics?*). Table 2 shows the descriptive values for students' interest and self-concept in physics in the pre- and post-test as well as the respective scale means transformed to POMP scores (Cohen et al., 1999). Scores are displayed separately for female and male participants.

The z-distributions of these two motivational variables were significantly negatively skewed. Skewness values were − 0.66 for dispositional interest, −0.68 for dispositional self-concept, −0.90 for situational interest, −0.73 for situational self-concept, and the standard error was estimated at 0.11 based on Tabachnick and Fidell (2013). This indicates a tendency toward a ceiling effect for the respective scales, which mirrors the study sample. Nevertheless, six *t*-tests were

TABLE 2 Descriptive values of Students' interest and self-concept in pre- and post-test

	All ($N = 509$)			Female ($n = 182$)			Male ($n = 327$)		
	M	SD	M%	M	SD	M%	M	SD	M%
Interest in physics									
Pre-test (dispositional)	4.29	0.94	66%	4.11	0.98	62%	4.39	0.90	68%
Post-test (situational)	4.84	0.84	77%	4.87	0.84	77%	4.82	0.83	76%
Physics self-concept									
Pre-test (dispositional)	4.32	1.00	66%	4.08	1.01	62%	4.46	0.97	69%
Post-test (situational)	4.70	0.77	74%	4.62	0.85	72%	4.75	0.73	75%

Abbreviations: M, scale mean value; SD, standard deviation; M%, mean value in percent of maximum possible score.

conducted, and the significance levels were interpreted following the Bonferroni correction with $m = 6$.

The dependent-samples t-test results confirmed that the mean value of students' specific (situational) interest triggered by the intervention was significantly higher than their general (dispositional) interest in physics before the intervention ($t(508) = -13.7, p < 0.001, r = 0.49$). This difference translates into a medium-sized positive effect with $d_{RM} = 0.6\,[0.5, 0.7]$.

Before the intervention in S'Cool LAB, female participants' interest in physics was significantly lower than that of male participants ($t(347) = 3.2, p = 0.002$), which quantifies into a small effect ($d = 0.3\,[0.1, 0.5]$); while there was no significant gender difference after the intervention ($t(507) = -0.6, p = 0.53$).

When comparing students' self-concept before and after the intervention, the results confirmed a significantly higher situational self-concept after the intervention ($t(508) = -8.6, p < 0.001, r = 0.38$) with an effect size at the Hattie hinge point ($d_{RM} = 0.4\,[0.2, 0.5]$). Moreover, there was a significant gender difference with respect to students' self-concept before the intervention ($t(507) = 4.2, p < 0.001$) with an effect size in favor of male participants again at the Hattie hinge point ($d = 0.4\,[0.2, 0.6]$); however, there was no significant gender difference after the intervention ($t(507) = 1.7, p = 0.10$).

4.2 | Effects of student and setting factors on motivational outcomes

In the following, we address Research Question 2 (*Which student and setting factors predict students' motivational outcomes?*). In particular, we report the results of single- and multilevel linear models with the motivational outcomes situational interest and self-concept as dependent variables and relevant student and setting factors as independent variables. In the multilevel models, students are treated as level-1 units and student groups as level-2 units. Variance components as well as VPCs and VIFs are reported in Table 3.

Students' situational interest varies greatly between groups with a VPC of 28% but variance partitioning is less pronounced for self-concept (VPC = 14%). Both VPCs exceeded 5% suggesting that multilevel modeling is necessary for this data set (Mehmetoglu & Jakobsen, 2016). Similarly, the VIFs (interest in physics: 6, physics self-concept: 4) indicate that the effective sample size would be severely overestimated when using single-level models.

4.2.1 | Situational interest

The evaluation of the assumptions of linear regression showed a deviation from a normal distribution and violation of homoscedasticity for the distribution of residuals. Also, three outliers with extremely standardized regression residuals $|z| > 3.29$ and 13 additional multivariate

TABLE 3 Student-level variance (σ_e^2) and group-level variance (σ_u^2) of motivational outcomes

	Student level		Group level			
	σ_e^2	SE	σ_u^2	SE	VPC	VIF
Situational interest in physics	0.74	0.05	0.29	0.09	28%	6
Situational physics self-concept	0.86	0.06	0.15	0.06	14%	4

Abbreviations: SE, standard error; VPC, variance partition coefficient; VIF, variance inflation factor.

outliers were removed from the sample. Nevertheless, variables were not transformed beyond the linear z-transformation, to maintain comparability between the different predictors. Furthermore, the estimation of fixed effects in multilevel modeling is robust against violation of normal distribution. Nevertheless, the standard errors for the regression coefficients can be slightly biased downwards particularly if the number of level-2 units is less than 50 (Maas & Hox, 2004). Therefore, a smaller alpha level may be appropriate when evaluating the significance of predictors. However, the conservative Bonferroni correction with $m = 13$ was already applied to account for the high number of predictors, which reduces the required alpha level significantly.

The results of both single- and two-level linear models for the outlier-adjusted sample ($N = 493$) are reported in Table 4. The size of significant regression coefficients is similar in both models for interest; p values are generally slightly larger in the two-level model, which reflects the smaller effective sample size caused by the clustering of the data. In the following, only the standardized regression coefficients of the two-level models are reported.

In both models, only four of the 13 predictors showed regression coefficients significantly different from zero with $p < 0.0004$ (Bonferroni correction with $m = 13$). The four predictors are the two student factors "dispositional interest" ($\beta_2 = 0.19^{***}$) and "curiosity state" ($\beta_2 = 0.21^{***}$), and the two setting factors "tinkering support" ($\beta_2 = 0.26^{***}$) and "tutor support" ($\beta_2 = 0.27^{***}$). In total, all 13 predictors accounted for 55% of the variance in the dependent variable in the single-level model for students' situational interest (adjusted $R^2 = 0.55, \text{SEE} = 0.64, F(13, 479) = 41, p < 0.001$), which indicates a good model fit according to Muijs (2004). Since adjusted R^2 penalizes the number of predictors, it is plausible that the value of 58% for analogue R^2 for the two-level model is slightly larger.

Moreover, adding student-level predictors significantly reduced the estimates of both variance components in the two-level model. In particular, the predictors reduced the student-level variance from 0.74 to 0.36 and the group-level variance from 0.29 to 0.07. Here, it comes as a surprise that student-level predictors have such an effect on reducing group-level variance. Indeed, the remaining proportion of group-level variance of 7% was reduced almost to the 5% multilevel modeling threshold suggested by Mehmetoglu and Jakobsen (2016).

4.2.2 | Situational self-concept

The evaluation of the assumptions of linear regression with respect to normality and homoscedasticity displayed very similar violations as described for the variable situational interest. Here, 14 cases were removed; five of them showed extreme residuals, additional nine cases were identified as multivariate outliers. Consequently, the outlier-adjusted sample ($N = 495$) was used for regression analysis. As in the model for situational interest, the size of significant regression coefficients and the amount of modeled variance were similar when comparing the results of the single- and two-level regression models, but the p values were slightly larger in the two-level model (Table 5). The Bonferroni correction was applied with $m = 12$.

Four regression coefficients were significantly different from zero with $p < 0.0004$. The strongest predictor of situational self-concept among the student factors was students' dispositional self-concept ($\beta_2 = 0.19^{***}$), which is consistent with the theoretical background of the self-concept construct. Among the setting factors, the strongest predictors were "cognitive load of the experiments" ($\beta_2 = -0.31^{***}$), "setting orientation due to preparation" ($\beta_2 = 0.23^{***}$), and "tutor support" ($\beta_2 = 0.19^{***}$). To conclude, all 12 predictors in the single-level

TABLE 4 Comparison of single- and two-level linear regression models of student and setting factors for participants' situational interest: Standardized regression coefficients and p values for the reduced sample without outliers ($N = 493$)

Predictor	Single-level model		Two-level model	
	β_1	p_1	β_2	p_2
(Constant)		0.60		0.85
Gender (female: 1, male: 0)	0.03	0.27	0.03	0.39
Age	−0.02	0.50	0.01	0.81
Dispositional interest in physics	0.21***	< 0.0004	0.19***	< 0.0004
Curiosity state particle physics	0.20***	< 0.0004	0.21***	< 0.0004
Experience hands-on experiments	−0.03	0.28	0.01	0.87
English skills (binary)	0.04	0.19	0.02	0.53
Experience out-of-school science	0.01	0.78	0.02	0.61
Experience particle physics	−0.04	0.27	−0.02	0.61
Tinkering support	0.29***	< 0.0004	0.26***	< 0.0004
Setting orientation	0.07	0.07	0.07	0.06
Lack of problems	−0.07	0.08	−0.05	0.13
Cognitive load experiments	−0.04	0.32	−0.07	0.07
Tutor support	0.27***	< 0.0004	0.27***	< 0.0004
Model fit	Adjusted $R^2 = 0.55$		Analog $R^2 = 0.58$	
student level variance σ_e^2 (SE)	–		0.35 (0.02)	
group level variance σ_u^2 (SE)	–		0.07 (0.03)	

Note: *** indicates $p < 0.0004$ (refers to an alpha level of 0.001 divided by $m = 13$).
Abbreviation: SE, standard error.

model accounted for 52% of the variance of students' situational self-concept ($R^2 = 0.52, \text{SEE} = 0.65, F(12,482) = 46, p < 0.001$), which indicates a good model fit (Muijs, 2004). As in the regression on situational interest, the value of 58% for analogue R^2 for the two-level model was slightly larger because it does not penalize the number of predictors. As before, adding student-level predictors significantly reduced the estimates of both variance components in the two-level model. The initial VPC of students' situational self-concept was 14%, only half as large as for situational interest (Table 3). Nevertheless, adding predictors significantly reduced both the student-level variance from 0.86 to 0.39 and the group-level variance from 0.15 to 0.03. The remaining proportion of group-level variance is almost negligible and lies below the 5% multilevel modeling threshold of Mehmetoglu and Jakobsen (2016).

4.2.3 | Variance components of student-level predictors

In the regression models, both for interest and self-concept, student-level predictors reduced group-level variance to a high extent. Thus, we decided to analyze the variance components of the predictors themselves, which complements the analysis of the variance components of the outcome variables presented in Table 3. In particular, we used multilevel modeling to

TABLE 5 Comparison of single- and two-level linear regression models of student and setting factors for participants' situational self-concept: Standardized regression coefficients and p values for the reduced sample without outliers ($N = 495$)

	Single-level model		Two-level model	
Predictor	β_1	p_1	β_2	p_2
(Constant)		0.02	−0.03	0.64
Gender (female: 1, male: 0)	0.00	0.96	−0.02	0.72
Age	0.02	0.47	0.02	0.62
Dispositional physics self-concept	0.21***	< 0.0004	0.19***	< 0.0004
Experience hands-on experiments	0.00	0.96	0.00	0.97
English skills (binary)	0.06	0.06	0.07	0.28
Experience out-of-school science	0.03	0.29	0.05	0.28
Experience particle physics	−0.02	0.50	−0.01	0.71
Tinkering support	0.10	0.04	0.08	0.07
Setting orientation	0.25***	< 0.0004	0.23***	< 0.0004
Lack of problems	0.02	0.57	0.03	0.39
Cognitive load experiments	−0.31***	< 0.0004	−0.31***	< 0.0004
Tutor support	0.16***	< 0.0004	0.19***	< 0.0004
Model fit	Adjusted $R^2 = 0.52$		Analog $R^2 = 0.58$	
student level variance σ_e^2 (SE)	–		0.39 (0.03)	
group level variance σ_u^2 (SE)	–		0.03 (0.01)	

Note: *** indicates $p < 0.0004$ (refers to an alpha level of 0.001 divided by $m = 12$).
Abbreviation: SE, standard error.

estimate by how much the presumed student-level predictors differed systematically between groups. Table 6 shows the variance components of all predictor variables that formed part of the regression models for the motivational outcomes interest or self-concept. To simplify the comparison and interpretation of predictors, all variables were z-transformed beforehand.

The highest VPCs were identified for students' experience with hands-on experiments (58%) and students' age (50%); that is, students participating within the same group shared very similar previous experiences with hands-on experimentation as well as were of similar age. Moreover, the student factor "dispositional interest" showed a large VPC of 35%, which implies that students within the same group shared a similarly low or high dispositional interest in physics before taking part in the intervention. Also, the VPCs of students' English skills (27%) and their previous experience with particle physics (28%) indicated significant similarities within student groups.

Across all 14 predictor variables, the lowest VPCs were identified for students' perceived lack of problems while doing experiments (5%), their previous experience with out-of-school learning (6%), as well as their perception of their cognitive load (10%) and support by tutors (10%). When comparing the two different types of factors, student factors showed much larger VPCs (average VPC of 27%) than setting factors (average VPC of 12%).

TABLE 6 Variance components of 14 predictors forming part of the regression models for motivational outcomes interest and self-concept

	Student-level		Group-level		
	σ_e^2	SE	σ_u^2	SE	VPC
Student factors					
Gender (female: 1, male: 0)	0.90	0.06	0.13	0.05	13%
Age	0.48	0.03	0.48	0.14	50%
Dispositional interest in physics	0.68	0.04	0.36	0.11	35%
Dispositional physics self-concept	0.87	0.06	0.13	0.05	13%
Curiosity state particle physics	0.85	0.05	0.18	0.07	17%
English skills (dichotomized)	0.75	0.05	0.28	0.09	27%
Experience hands-on experiments	0.42	0.03	0.58	0.17	58%
Experience out-of-school learning	0.94	0.06	0.06	0.03	6%
Experience with particle physics	0.75	0.05	0.29	0.09	28%
Setting factors					
Tinkering support	0.85	0.05	0.17	0.06	17%
Setting orientation	0.83	0.05	0.19	0.07	19%
Lack of problems	0.95	0.06	0.05	0.03	5%
Cognitive load experiments	0.89	0.06	0.10	0.04	10%
Perceived tutor support	0.92	0.06	0.10	0.04	10%

Abbreviations: σ_e^2, student-level variance; σ_u^2, group-level variance; SE, standard error; VPC, variance partition coefficient.

5 | DISCUSSION

5.1 | General motivational outcomes interest and self-concept (RQ1)

The intervention at S'Cool LAB triggered very high situational interest (POMP score: 77%) and self-concept (POMP score: 74%), exceeding values reported by other studies at science outreach labs such as Priemer et al. (2018) (POMP score: 59%), Itzek-Greulich et al. (2017) (POMP score: 56%), and Pawek (2009) (POMP score: 70%). Moreover, students' situational interest triggered by the intervention was significantly higher than their dispositional interest in physics in general ($d_{RM} = 0.6$). An analogous result was found for students' self-concept in physics ($d_{RM} = 0.4$). Considering the relatively short intervention time (4.5 h) and the fact that improvement of students' self-concept was not the main focus of the intervention, this finding is promising and in the range of previous meta-analytic findings (d = 0.51 [0.13]) for the impact of self-concept interventions (O'Mara et al., 2006). Specifically, these results were obtained for a science outreach lab at a large, international research facility for particle physics. They thus provide, to the best of our knowledge, the first empirical evidence for the hypothetical motivational potential of experiencing "big science" at an out-of-school learning opportunity, as put forward by Braund and Reiss (2006).

Prior to the intervention, girls reported lower physics-related interest (d = 0.3) and self-concept (d = 0.4) than boys. However, after the intervention, there were no significant gender differences. Instead, girls and boys showed similarly high triggered interest and self-concept. Consequently, the intervention helped to close the gender gap (Baram-Tsabari & Yarden, 2011) with regard to these variables. Previous studies report conflicting evidence with respect to closing of a motivational gender gap in science. On the one hand, a similar positive trend was reported by studies at other science outreach labs (Itzek-Greulich & Vollmer, 2017; Mokhonko et al., 2014; Pawek, 2009) and by studies investigating science camps (Levine et al., 2015; Wang & Frye, 2019). On the other hand, physics instruction in regular settings has been found to have negative effects on female learners' interest and self-concept at secondary level I (Häussler & Hoffmann, 2002). More recently, similar findings were reported at university level: Li and Singh (2021) found that a lecture-based physics learning environment even increased gender gaps in the motivational constructs self-efficacy and interest. Moreover, Nissen and Shemwell (2016) report a similar detrimental effect of interactive-engagement physics courses for female learners' self-efficacy. One possible interpretation of the results of our study in this context is that strong stereotypes about both physics and of physics-related gender roles, originating from experiences in everyday life and ordinary educational settings and (Kessels et al., 2006; OECD, 2015; Ramsey et al., 2013; Taconis & Kessels, 2009), become less influential in a new and exploratory setting like that of a science outreach lab (Euler, 2005; Hannover & Kessels, 2002). In summary, our study extents the encouraging findings about the potential of science outreach labs with respect to the motivational gender gap by demonstrating a positive effect on high-school students from a broad range of 13 countries.

It has to be noted that participants of the science outreach lab in this study were generally students whose dispositional interest and self-concept in physics were already high before the intervention. On the one hand, this means that the intervention had a significant and positive effect even for these students; on the other hand, there is no guarantee that a similar effect would be observed for other student groups. However, further information about this issue can be gained by the analysis of the predictors taken into account, to which we turn now.

5.2 | Relevant predictors of motivational outcomes (RQ2)

5.2.1 | The role of student factors

Higher levels of situational interest were reached by students with a high level of initial interest, in line with previous research on interest in general (for example, Rotgans & Schmidt, 2018) and interest at science outreach labs in particular (e.g., Itzek-Greulich & Vollmer, 2017). Moreover, students who were already curious about particle physics before the intervention demonstrated higher levels of situational interest after taking part in particle physics hands-on activities, which can be explained by the close relationship between epistemic curiosity with the construct of interest (Alexander & Grossnickle, 2016). Similarly, higher levels of situational self-concept were reached by students who had high physics self-concept.

Furthermore, the results of the regression models suggest that the motivational outcomes of the intervention were independent of all student factors capturing potentially relevant prior experience and knowledge such as age, experience with particle physics, hands-on experiments, or out-of-school science learning. Moreover, even students' English skills did not have a significant influence, which is compatible with the results of a study at a bilingual (English and German) science outreach lab showing no effect of language on learning (Rodenhauser, 2018; Rodenhauser & Preisfeld, 2018).

These findings provide another aspect of the promising potential of out-of-school learning opportunities, especially if they are catering to a diverse audience from different countries and educational backgrounds. Indeed, European students aged 16–19 were able to follow and benefit from hands-on activities in English, even if English was not their mother tongue, and even if they had little prior experience with hands-on experimentation, out-of-school learning, or the learning topic particle physics.

5.2.2 | The role of setting factors in predicting interest

Among the setting 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. Hence, even after controlling for student factors such as their prior experiences and interest, students' perception of support—both by the learning environment and the educators in the room—was a crucial element of their science outreach lab experience. The results of this quantitative study thus complement the qualitative study results by Habig et al. (2020) who describe the interaction with STEM professionals as important design principle of successful out-of-school science learning programs. Moreover, this result is consistent with findings by Glowinski and Bayrhuber (2011), who identified support by scientists ("quality of instruction") as important predictor of situational interest, as well as with findings by Pawek (2009) who reports support by tutors and a positive learning atmosphere as strong predictors of situational interest. Consequently, this study provides further evidence across learners from 13 countries that supportive interaction with educators is a crucial success factor of out-of-school learning opportunities.

5.2.3 | The role of setting factors in predicting self-concept

With respect to the motivational outcome "situational self-concept" students reporting high levels of setting orientation due to their preparation (e.g., students knew what to expect and were familiar with the schedule) and low levels of cognitive load (e.g., students did not find the experiments difficult or the instructions hard to understand) benefited more from the intervention. As for interest, the perception of support by tutors was another strong predictor. Hence, even after controlling for student factors such as their prior experiences and dispositional self-concept in physics, students' perception of their cognitive preparedness, their cognitive load as well as the support they received were vital components of their science outreach lab experience. This is consistent with other research emphasizing the importance of novelty-reducing preparation (i.e., setting orientation) and the impact of cognitive load on learning and development. For example, Streller (2015) confirmed positive effects on students' situational self-concept (and interest) when preparing them via an online portal. Also, Molz (2016) confirmed that preparation aiming at reducing setting novelty had a positive effect on the motivation component "self-concept". Potential detrimental effects of a lack of familiarity and overwhelming novelty have been mentioned by other researchers investigating science outreach labs (Dairianathan & Subramaniam, 2011; Itzek-Greulich et al., 2015; Randler et al., 2005). However, only a few studies systematically accounted for cognitive load at out-of-school science learning offers (Scharfenberg & Bogner, 2010; Goldschmidt et al., 2016; Mierdel & Bogner, 2021; Röllke et al., 2020; Van Winkle, 2012). The present work adds to this research by highlighting cognitive load as the strongest predictor of self-concept. This result suggests that cognitive load should be included systematically in research about motivational outcomes of science outreach labs. Moreover, the specific setting of our study allowed to replicate previous findings from almost exclusively German science outreach labs focusing on life science with a sample of students from a broad range of countries and a science outreach lab in the area of physics.

The differences between the regression coefficients of setting factors when modeling interest and self-concept might be explained by the different roles of affective and cognitive aspects with respect to these two motivational outcomes. Although interest involves both affective and cognitive components, the latter is more important when interest develops into a disposition over time (Hidi, 2006), while for situational interest, the affective component might be dominant. Consequently, it seems plausible that not the cognitive setting factors but rather the social and affective factors are relevant predictors of students' situational interest. In contrast, self-concept, defined as a collection of cognitive concepts about oneself (Gutman & Schoon, 2013; Valentine et al., 2004), does not have prominent affective components. Thus, it is plausible that cognitive setting factors are the most relevant student-level predictors of situational self-concept. For both motivational outcomes, however, the perception of tutors stands out as crucial laboratory characteristic.

5.2.4 | Overall effects of student and setting factors

We hypothesized that the motivational outcomes of science outreach labs strongly depend on student factors and students' perception of setting factors. In our study, 58% of the variance of students' situational interest as well as their self-concept was accounted for by predictors in the multilevel models. This good model fit (Muijs, 2004) suggests that our strategy of modeling motivational outcomes on research-based and well-operationalized factors (9 student factors and 5 setting factors) proved successful. Indeed, by quantitatively assessing and comparing a high number of student and setting factors across multiple countries, we present a novel synthesis of these impact factors.

5.3 | Student- and group-level influences on situational interest and self-concept

Not only did the motivational outcomes interest and self-concept show a significant proportion of group-level variance, but also many of the student and setting factors that served as predictors in the multilevel models of motivational outcomes. However, students in the same group always came from the same school and often shared the same physics teacher. Hence, the teachers as well as the conditions at school might affect certain student factors.

Indeed, the VPC of students' experience with hands-on experiments (58%) confirms a large group-level influence, consistent with what one would expect for groups of students taught by different teachers under different education systems. Similarly, students' dispositional interest, English skills, and previous experience with particle physics all demonstrated VPCs above 27%, which can also be attributed to different teachers and education systems. However, students' curiosity toward particle physics showed a considerably smaller VPC of 17%. This comes as a surprise because epistemic curiosity and interest are closely related constructs, and both should be affected by group-level variables such as teachers in a similar way. This could indicate that students' curiosity toward particle physics is more independent of prior teaching influences. Similarly, students' dispositional self-concept demonstrated a comparatively small VPC of 13%. This is consistent with the VPC of students' situational self-concept (14%), suggesting that self-concept is less affected by group-level variables, such as teachers, and acts rather as an individual student characteristic. Surprisingly, students' previous experience with out-of-school learning showed one of the smallest VPCs of 6%. This suggests that a large proportion of students' out-of-school science learning experience was not acquired during school trips. Instead, students might have acquired the majority of their out-of-school science learning experience outside of their school group, for example, during family trips or extracurricular activities.

Interestingly, all setting factors showed relatively low proportions of group-level variance. For example, only 10% of the variance of students' perception of support by tutors was estimated at the group level. However, tutors varied between groups and it would therefore be plausible if this variable had a larger group-level variance. Although, students' perceptions of tutors were extremely positive, and the variables' variance was quite small in the first place, the differences between tutors were not that important to students. The 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 the science outreach lab, it makes sense that students within one group perceived their setting orientation in a similar way to some degree.

When comparing the two different groups of predictors, student factors showed much larger VPCs (average VPC of 27%) than setting factors (average VPC of 12%). Here, student factors refer to student characteristics while setting factors characterize students' perception of laboratory characteristics. Hence, it comes as no surprise that the clustering of students in groups affects variance components of student characteristics more than variance components of perceived laboratory characteristics, even if student factors affect setting factors to a certain degree.

Note that student-level predictors accounted for group-level variance because the predictors themselves showed a substantial proportion of group-level variance. In particular, student factors such as students' dispositional interest did not act like "pure" student-level predictors for the outcome "situational interest," but instead were a combination of student- and group-level predictors. Indeed, the results suggest that the large VPC for the motivational outcome "situational interest" can be traced back, at least partly, to the large VPCs of the predictors. This suggests that a significant proportion of variance in the outcomes of this particular out-of-school learning opportunity cannot be attributed to students' perception of the learning activities but needs to be attributed to the initial conditions of the group.

In summary, the present results provide a synthesis of 14 research-based predictors on motivational outcomes of science outreach labs hitherto unavailable. In particular, a group of nine student factors capturing characteristics of learners was combined for the first time with a group of five setting factors. First, this allowed to draw attention to the importance of specific factors such as tutor support, cognitive load, or prior setting orientations, which should be more systematically be taken into account in future motivational research about science outreach labs. Second, our results allow to rank student and setting factors suggested by previous research based on their importance in predicting motivational outcomes. For example, neither prior experience or knowledge nor English language skills were significant predictors. However, almost all setting factors, in particular, all forms of support (prior orientation, tinkering, tutors), and cognitive load were highlighted as very important predictors even after controlling for student factors. Third, our results allow to interpret within the multi-level analysis, which of the predictors have larger or smaller variance components on the group level. Thus, we draw attention to the benefits of multi-level analysis in out-of-school science learning settings as it provides insights into differences between the initial conditions of groups of learners, which can even be used to better adapt learning activities to the needs of specific groups. Finally, our models account for almost 60% of the variance of both motivational outcomes confirming a suitable selection and a good model fit of the research-based impact factors in this study.

6 | CONCLUSIONS

6.1 | Summary and implications

Previous studies had suggested that science outreach labs have a unique educational potential based on their unique features, such as access to rare equipment, hands-on experimentation, and contact with scientists. However, the effectiveness of these educational opportunities depends on their design and how they succeed in realizing their potential in the actual interaction with students. 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. Consequently, the present study was designed to explore a large set of research-based impact factors of hands-on learning activities at science outreach labs.

The findings of this study contribute in several ways to our understanding of motivational effects at science outreach labs: First, the results confirm a considerable motivational effectiveness of the intervention based on particle physics experiments at a large research facility. Despite the short intervention time of only 4.5 h, participation led to very high situational interest and situational self-concept in physics. Note that this is, to the best of our knowledge, the first empirical evidence for the hypothetical motivational potential of experiencing "big science" at an out-of-school learning opportunity (Braund & Reiss, 2006). Moreover, these findings were observed consistently on a sample consisting of students from 13 (mostly European) countries, thus considerably extending previous research on science outreach labs that focused mainly on a few selected countries.

Second, we observe a closing of a motivational gender gap (Baram-Tsabari & Yarden, 2011) with respect to science. Although female participants reported a slightly lower physics-related interest and self-concept than male participants, they showed larger gains for both dimensions, and no gender differences were identified after the intervention. Here, two aspects add to and expand our current understanding of the gender gap in interest and self-concept in physics. This study confirmed an initial gender gap in students' physics-related interest and self-concept even for high-achieving students taking part predominately in advanced-level physics courses. Moreover, both the initial gender gap and its closing at a science outreach lab were observed consistently across 13 (mostly European) countries with different educational systems.

Third, to advance our understanding of factors impacting the learning experience in an out-of-school setting, we followed the recommendation by Falk and Storksdieck (2005) to study homogeneous groups of learners with a large sample size while using a careful selection of well-established predictors in multilevel models to account for students' different initial conditions as well as their perceptions of relevant setting factors. Our research contributes to existing knowledge in the field about that matter in several ways. (i) We provide a new synthesis of 14 predictors suggested by previous research, allowing to identify predictors among both student and setting factors and rank them according to their influence. Together, the chosen predictors accounted for almost 60% of the variance in students' motivational outcomes, exceeding other recent studies by a factor of two. We interpret this as a step toward a systematic measurement and understanding, providing valuable information about science outreach labs and how to optimize their effectiveness. (ii) None of the learner characteristics regarding potentially relevant prior experience and knowledge (particle physics, hands-on experiments, out-of-school science learning, and English) was found to have a significant influence. This is particularly interesting when considering the encouraging motivational effects of the intervention and the closing of the gender gap. Taken together, and in view of the broad sample of countries in the sample, our results add evidence that science outreach labs can provide a motivational enhancement for diverse kinds of learners, of different gender, prior experience and knowledge, and from different educational backgrounds. (iii) On the other hand, we provide empirical evidence for predictors that *are* influential for the motivational enhancement at science outreach labs and that can help inform researchers and practitioners about relevant success factors. We find that students' perception of support both by the learning environment and the educators in the room was a crucial settings factor with respect to their triggered situational interest. Furthermore, students' perception of their cognitive preparedness and their cognitive load were vital setting factors with respect to their situational self-concept. These results suggest that, even after controlling for a number of student factors, guidance, scaffolding, and preparation are highly valuable elements of science outreach labs even for high-achieving students and when targeting affective outcomes.

Fourth, multi-level analyses provided informative insights into the variance components not just of the dependent variables but also of the independent variables, which allowed to attribute effects of predictors to individual as well as group differences. These results can also allow to adapt learning activities to the specific needs of groups of learners. Moreover, the analysis of variance partition coefficients for the 14 student and setting factors provides evidence for the degree of necessity of multi-level analyses when investigating science outreach labs. Indeed, our findings strongly support the use of multilevel models on the clustered data that are usually produced at out-of-school learning opportunities as it avoids overestimating effective sample sizes.

Fifth, we also present a set of valid, reliable, and short measures that might support future research in informal learning settings.

6.2 | Limitations and strengths

There are several limitations to this study. First, the correlational research design does not allow to draw conclusions about the causal relationships among the measured variables. Instead, this study assesses the relationships between motivational variables as dependent variables and a set of predictor variables. For instance, questionnaire items in the post-test specifically referred to the effects caused by the particle physics experiments in S'Cool LAB. Nevertheless, students' answers in the post-test might have been influenced by factors other than those related to the intervention. For example, participation in the guided tour at CERN before the intervention in S'Cool LAB or follow-up activities prepared by the teacher between the intervention in S'Cool LAB and the post-test might also have affected the educational outcomes reported by the students within this study. Moreover, findings of the reduced motivational gender gap are *consistent* with the interpretation that science outreach labs provide a more neutral learning environment in which certain gender stereotypes might not be activated. However, more research would be needed to actually *establish* this interpretation. Similarly, future research should compare the motivational effects of S'Cool LAB at CERN with other opportunities to experience "big science".

Second, the intervention of this study focused on particle physics experiments. Here, students might find particle physics generally interesting, or more interesting than other physics or science topics. Consequently, comparisons of absolute scores for interest and self-concept should be interpreted with caution. Moreover, future studies should investigate the role of student-level and setting-level predictors in a science outreach lab focusing on a different topic.

Third, as is usually the case with out-of-school learning opportunities, this study relied on the research sample provided by participants taking part in the offers of the presented out-of-school learning lab. Therefore, students were not randomly selected. Instead, groups of students were selected based on their teacher's motivation. Many of the groups consisted of students studying advanced level physics courses. Consequently, students in this study exhibited a high dispositional interest in physics as well as a high dispositional physics self-concept. This bias (via ceiling effects) is also reflected in the deviations from a normal distribution of the regression residuals of linear models. Also, this study sample showed a gender bias with only about one-third female participants. Although the use of online questionnaires proved successful, only about two-thirds of the students who filled out the pre-test also filled out the post-test, which might have resulted in an additional bias in the sample toward students who had a positive experience. Although the results of this study are consistent with previous research studies, they may not be readily generalizable to other populations of high-school students. However, the very positive motivational impact of the intervention was not significantly influenced by relevant prior experience and knowledge. Consequently, it is justified to expect that the intervention might also be beneficial for average-achieving students (potentially even leading to higher effect sizes, as there the present data showed a tendency toward a ceiling effect for the highly motivated students of our sample). Nevertheless, future research should investigate not only the effect of this intervention but also the relevance of the identified predictors for a different sample. A more heterogeneous sample can be recruited by inviting, for example, randomly selected school groups of local public schools.

Fourth, while the set or predictor variables considered here is broader than in previous studies at science outreach labs, there is of course no claim to be exhaustive. For instance, other research has focused on personality traits like conscientiousness (Itzek-Greulich & Vollmer, 2017). Combining results from different studies in a systematic way, as proposed in the present work, may lead to an increasingly comprehensive understanding of motivational effects at science outreach labs.

Furthermore, the results suggest that cognitive motivation components are affected more by cognitive factors, whereas affective motivation components are affected more by affective and social factors. Future research could examine this relationship more in detail, for example, through manipulating specific novelty types and investigating the effect on different types of affective and cognitive outcomes.

Finally, this study focused on effects on situational interest and self-concept and did not study effects on the more stable dispositional variables. Hence, our study does not allow to elucidate the complex "trait/state" interactions discussed in the literature (see, e.g., Ainley, 2017, Knogler et al., 2015; Krapp et al., 1992), which would need more detailed research. In particular, effects of single short-time interventions are usually not large and fade within a few months (Jarvis & Pell, 2005). Nevertheless, future studies should explore long-term effects of interventions at science outreach labs and ways to enhance them by appropriate instructional measures such as follow-up activities.

The strength of the presented study lies in the combination of an intervention with a unique educational potential at a large research facility, a large multi-national student sample, and a multilevel modeling approach. Furthermore, the research-based assessment of a broad selection of student and setting factors allowed to systematically derive a novel synthesis of impact factors that affect the educational effectiveness of science outreach labs. In summary, we consider that the science outreach lab S'Cool LAB at CERN appears as a promising approach to provide an appreciable boost for physics-related interest and self-concept of high-school students. Further analyses show that this holds for diverse kinds of learners, of different gender, prior experience and knowledge, and from different educational backgrounds. Moreover, they allow to identify essential influence factors on the student and setting level which are of interest for future research about and the practical design of science outreach lab interventions alike.

ACKNOWLEDGMENT

This article is based on the doctoral dissertation of the first author, who was sponsored by the German Federal Ministry of Education and Research through a Wolfgang-Gentner Scholarship. Open access funding provided by European Organization for Nuclear Research.

ORCID

Julia Woithe https://orcid.org/0000-0002-4772-2989

ENDNOTES

[1] We use the term "out-of-school learning" and not "informal learning," because these learning opportunities can be closely connected to formal learning in school (Eshach, 2007).

[2] A part of the literature cited below refers to museums. However, some features of learning settings at science outreach labs differ from those in museums; for instance, students' choice is more limited during the hands-on activities at science outreach labs. Nevertheless, similar factors have been shown to influence the motivational effectiveness of hands-on sessions in both settings.

[3] Note that Hattie (2008) points out that this is an element of discussion to be used with circumspection, not a threshold value to be blindly applied. Lower effect sizes might well be worth consideration, depending on available alternatives, effort, and so on, and vice versa for higher effect sizes.

REFERENCES

Abrahams, I. (2009). Does practical work really motivate? A study of the affective value of practical work in secondary school science. *International Journal of Science Education*, *31*(17), 2335–2353. https://doi.org/10.1080/09500690802342836

Ainley, M. (2017). Interest: Knowns, unknowns, and basic processes. In P. A. O'Keefe & J. M. Harackiewicz (Eds.), *The science of interest*. Springer.

Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, *94*(3), 545–561. https://doi.org/10.1037/0022-0663.94.3.545

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), 411–428. https://doi.org/10.1016/S0959-4752(01)00008-1

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 ed., pp. 188–208). Routledge.

Andre, T., & Windschitl, M. (2003). Interest, epistemological belief, and intentional conceptual change. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 173–197). Lawrence Erlbaum.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Baram-Tsabari, A., & Yarden, A. (2011). Quantifying the gender gap in science interests. *International Journal of Science and Mathematics Education*, *9*(3), 523–550. https://doi.org/10.1007/s10763-010-9194-7

Bettelheim, B. (1980). Curiosity—Its applicability in a museum setting. In Office of Museum Programs (Ed.), *Proceedings of the Children in Museums International Symposium* (pp. 22–31). Smithsonian Institution.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century century skills*. Springer. https://doi.org/10.1007/978-94-007-2324-5_2

Boiko, A., Nistor, A., Kudenko, I., & Gras-Velazquez, A. (2019). *The attractiveness of science, technology, engineering and mathematics subjects, Results from five countries*. European Schoolnet.

Bong, M., Lee, S. K., & Woo, Y. (2015). The roles of interest and self-efficacy in the decision to pursue mathematics and science. In K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest, the self, and K-16 mathematics and science learning*. American Educational Research Association.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Börlin, J. (2012). *Das experiment als Lerngelegenheit: Vom interkulturellen Vergleich des Physikunterrichts zu Merkmalen seiner Qualität [the experiment as a learning opportunity: Intercultural comparison of physics lessons via characteristics of their quality]* (p. 132). Logos Verlag.

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), 1373–1388. https://doi.org/10.1080/09500690500498419

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), 297–331. https://doi.org/10.3102/00028312034002297

CERN. (2020). S'Cool LAB. Retrieved from https://scoollab.web.cern.ch/.

Chi, B., Dorph, R., & Reisman, L. (2015). *Evidence & impact: Museum-managed STEM programs in out-of-school settings. National Research Council Committee on successful out-of-school STEM*. National Research Council.

Cinici, A., & Demir, Y. (2013). Teaching through cooperative POE tasks: A path to conceptual change. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *86*(1), 1–10. https://doi.org/10.1080/00098655.2012.712557

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). 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), 315–346. https://doi.org/10.1207/S15327906MBR3403_2

Corrigan, D., Buntting, C., Jones, A., & Loughran, J. (2018). *Navigating the changing landscape of formal and informal science learning opportunities*. Springer. https://doi.org/10.1007/978-3-319-89761-5_1

Cors, R. (2016). *Informal Science Learning: An investigation of how novelty and motivation affect interest development at a mobile laboratory [Doctoral dissertation]*. University of Geneva. https://doi.org/10.13097/archive-ouverte/unige:91515

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 (PriSE)*, *1*(1), 1–19. https://doi.org/10.25321/prise.2018.521

Dairianathan, A., & Subramaniam, R. (2011). Learning about inheritance in an out-of-school setting. *International Journal of Science Education*, *33*(8), 1079–1108. https://doi.org/10.1080/09500693.2010.494185

Doran, R. L. (1980). *Basic measurement and evaluation of science instruction*. National Science Teachers Association.

Dorfman, B. S., & Fortus, D. (2019). Students' self-efficacy for science in different school systems. *Journal of Research in Science Teaching*, *56*(8), 1037–1059. https://doi.org/10.1002/tea.21542

Durik, A. M., Hulleman, C. S., & Harackiewicz, J. M. (2015). One size fits some: Instructional enhancements to promote interest don't work the same for everyone. In K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest, the self, and K-16 mathematics and science learning*. American Educational Research Association.

Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology*, *16*(2), 171–190. https://doi.org/10.1007/s10956-006-9027-1

Euler, M. (2005). Hands-on science and informal learning: Challenge and potentials of authentic lab activities. In G. Planinšič & A. Mohorič (Eds.), *Informal learning and public understanding of physics, Book of Selected Contributions*. GIREP.

Evers, A., Hagemeister, C., Høstmælingen, A., Lindley, P., Muñiz, J., & Sjöberg, A. (2013). *EFPA review model for the description and evaluation of psychological and educational tests*. European Federation of Psychology Associations.

Exploratorium. (2020). Retrieved from https://www.exploratorium.edu/.

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*, *41*(8), 835–860. https://doi.org/10.1002/tea.20026

Falk, J., & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. Alta Mira Press.

Falk, J., & Storksdieck, M. (2005). Using the contextual model of learning to understand visitor learning from a science center exhibition. *Science Education*, *89*(5), 744–778. https://doi.org/10.1002/sce.20078

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), 127–134. https://doi.org/10.1002/tea.3660150207

Fermilab. (2020). Lederman Science Center. Retrieved from https://ed.fnal.gov/lsc/.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*(C), 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Friedman, A. J. (2010). The evolution of the science museum. *Physics Today*, *63*(10), 45–51. https://doi.org/10.1063/1.3502548

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), 371–392.

Goff, E. E., Mulvey, K. L., Irvin, M. J., & Hartstone-Rose, A. (2020). The effects of prior informal science and math experiences on undergraduate STEM identity. *Research in Science & Technological Education*, *38*(3), 272–288. https://doi.org/10.1080/02635143.2019.1627307

Goldschmidt, M., & Bogner, F. X. (2016). Learning about genetic engineering in an outreach laboratory: Influence of motivation and gender on Students' cognitive achievement. *International Journal of Science Education, Part B*, *6*(2), 166–187. https://doi.org/10.1080/21548455.2015.1031293

Goldschmidt, M., Scharfenberg, F. J., & Bogner, F. X. (2016). Instructional efficiency of different discussion approaches in an outreach laboratory: Teacher-guided versus student-centered. *The Journal of Educational Research*, *109*(1), 27–36. https://doi.org/10.1080/00220671.2014.917601

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics*, *1*(4), 223–231. https://doi.org/10.1207/S15328031US0104_02

Gunstone, R. F., & Champagne, A. B. (1990). Promoting conceptual change in the laboratory. In E. Hegarty-Hazel (Ed.), *The student laboratory and the science curriculum* (pp. 159–182). Routledge.

Gutman, L. M., & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people*. Education Endowment Foundation.

Habig, B., Gupta, P., Levine, B., & Adams, J. (2020). An informal science education program's impact on STEM major and STEM career outcomes. *Research in Science Education*, *50*(3), 1051–1074. https://doi.org/10.1007/s11165-018-9722-y

Hannover, B., & Kessels, U. (2002). Challenge the science-stereotype. Der Einfluss von Technik-Freizeitkursen auf das Naturwissenschaften-Stereotyp von Schülerinnen und Schülern. In M. Prenzel & J. Doll (Eds.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (pp. 341–358). Beltz.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, *72*(358), 320–338. https://doi.org/10.1080/01621459.1977.10480998

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hausamann, D. (2012). Extracurricular science labs for STEM talent support. *Roeper Review*, *34*(3), 170–182. https://doi.org/10.1080/02783193.2012.686424

Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of Research in Science Teaching*, *39*(9), 870–888. https://doi.org/10.1002/tea.10048

Heureka (2020). School programs. Retrieved from https://www.heureka.fi/kouluohjelmat-2/?lang=en.

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review*, *1*(2), 69–82. https://doi.org/10.1016/j.edurev.2006.09.001

Hidi, S., & Baird, W. (1986). Interestingness-a neglected variable in discourse processing. *Cognitive Science*, *10*(2), 179–194. https://doi.org/10.1207/s15516709cog1002_3

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hirth, M. (2019). *Akustische Untersuchungen mit dem smartphone und tablet-Computern—Fachliche und didaktische Aspekte [acoustic investigations using smartphones and tablets—Scientific and educational aspects] [doctoral dissertation, Technische Universität Kaiserslautern]*. Verlag Dr. Hut.

Hochberg, K. (2016). iMechanics: Smartphones als Experimentiermittel im Physikunterricht der Sekundarstufe II [iMechanics: smartphones as experimental tools in secondary II physics lessons] [Doctoral dissertation, Technische Universität Kaiserslautern]. URN: urn:nbn:de:hbz:386-kluedo-44454.

Hochberg, K., Kuhn, J., & Müller, A. (2018). Using smartphones as experimental tools—Effects on interest, curiosity, and learning in physics education. *Journal of Science Education and Technology*, *27*(5), 385–403. https://doi.org/10.1007/s10956-018-9731-7

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. John Wiley & Sons, Inc.

Hodson, D. (1993). Re-thinking old ways: Towards a more critical approach to practical work in school science. *Studies in Science Education*, *22*(1), 85–142. https://doi.org/10.1080/03057269308560022

Hoffmann, L., Häußler, P., & Peters-Haft, S. (1997). *An den Interessen von Mädchen und Jungen orientierter Physikunterricht [physics lessons oriented on girls' and boys' interests]*. IPN.

Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, *88*(1), 28–54. https://doi.org/10.1002/sce.10106

Hofstein, A., & Mamlok-Naaman, R. (2007). Special issue on experiments and the laboratory in chemistry education. *Chemistry Education Research and Practice*, *8*(2), 105–273.

Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, *42*, 791–806. https://doi.org/10.1002/tea.20072

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Huang, F. L. (2018). Multilevel modeling myths. *School Psychology Quarterly*, *33*(3), 492–499. https://doi.org/10.1037/spq0000272

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*, 43–52. https://doi.org/10.1016/j.learninstruc.2015.03.003

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*, 98–115. https://doi.org/10.1016/j.cedpsych.2016.09.005

Itzek-Greulich, H., & Vollmer, C. (2017). Emotional and motivational outcomes of lab work in the secondary intermediate track: The contribution of a science center outreach lab. *Journal of Research in Science Teaching*, *54*(1), 3–28. https://doi.org/10.1002/tea.21334

Jarvis, T., & Pell, A. (2005). Factors influencing elementary school children's attitudes toward science before, during, and after a visit to the UKNational Space Centre. *Journal of Research in Science Teaching*, *42*(1), 53–83. https://doi.org/10.1002/tea.20045

Kessels, U., Rau, M., & Hannover, B. (2006). What goes well with physics? Measuring and altering the image of science. *British Journal of Educational Psychology*, *76*(4), 761–780. https://doi.org/10.1348/000709905X59961

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), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Knogler, M., Harackiewicz, J. M., Gegenfurtner, A., & Lewalter, D. (2015). How situational is situational interest? Investigating the longitudinal structure of situational interest. *Contemporary Educational Psychology*, *43*, 39–50. https://doi.org/10.1016/j.cedpsych.2015.08.004

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*(4), 383–409. https://doi.org/10.1016/S0959-4752(01)00011-1

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning, and development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3–25). Erlbaum.

Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, *33*(1), 27–50. https://doi.org/10.1080/09500693.2010.518645

Kuhn, J., & Müller, A. (2014). Context-based science education by newspaper story problems: A study on motivation and learning effects. *Perspectives in Science*, *2*(1–4), 5–21.

Lernort Labor (2020). Schülerlabor Atlas. Retrieved from https://www.schuelerlabor-atlas.de/home/LabListe.

Levine, M., Serio, N., Radaram, B., Chaudhuri, S., & Talbert, W. (2015). Addressing the STEM gender gap by designing and implementing an educational outreach chemistry camp for middle school girls. *Journal of Chemical Education*, *92*(10), 1639–1644. https://doi.org/10.1021/ed500945g

Li, Y., & Singh, C. (2021). Effect of gender, self-efficacy, and interest on perception of the learning environment and outcomes in calculus-based introductory physics courses. *Physical Review Physics Education Research*, *17*(1), 010143.

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment*, *80*(1), 75–86. https://doi.org/10.1207/S15327752JPA8001_16

Lunetta, V. N., Hofstein, A., & Clough, M. P. (2007). Learning and teaching in the school science laboratory: An analysis of research, theory, and practice. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 393–441). Lawrence Erlbaum.

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 [Commissioned report].

Maas, C. J. M., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis*, *46*(3), 427–440. https://doi.org/10.1016/j.csda.2003.08.006

Mahalanobis, P. C. (1936). *On the generalized distance in statistics*. National Institute of Science.

Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). STEM: Country comparisons: International comparisons of science, technology, engineering and mathematics (STEM) education [Final report]. Retrieved from http://hdl.handle.net/10536/DRO/DU:30059041

Marsh, H., & 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), 542–552. https://doi.org/10.1177/0146167207312313

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*(4), 623–636. https://doi.org/10.1037/0022-0663.82.4.623

Marsh, H. W., Martin, A. J., Yeung, A. S., & Craven, R. G. (2017). Competence self-perceptions. In A. J. Elliot, C. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (pp. 85–115). The Guilford Press.

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), 291–309. https://doi.org/10.1037/0022-0663.100.2.291

Mehmetoglu, M., & Jakobsen, T. G. (2016). *Applied statistics using Stata: A guide for the social sciences*. Sage Publications.

Mierdel, J., & Bogner, F. X. (2021). Investigations of Modellers and model viewers in an out-of-school Gene Technology laboratory. *Research in Science Education*, *51*, 801–822. https://doi.org/10.1007/s11165-019-09871-3

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. https://doi.org/10.1103/PhysRevSTPER.9.020113

Mokhonko, S., Nickolaus, R., & Windaus, A. (2014). Förderung von Mädchen in Naturwissenschaften: Schülerlabore und ihre Effekte. *Zeitschrift für Didaktik der Naturwissenschaften*, *20*(1), 143–159. https://doi.org/10.1007/s40573-014-0016-2

Molz, A. (2016). *Verbindung von Schülerlabor und Schulunterricht-Auswirkungen auf motivation und Kognition im Fach Physik [linking a science outreach lab and school lessons—Effects on motivation and cognition in physics] [doctoral dissertation, Technische Universität Kaiserslautern]*. Verlag Dr. Hut.

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), 105–125. https://doi.org/10.1037/1082-989x.7.1.105

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. Sage Publications.

Mujtaba, T., & Reiss, M. J. (2013). Inequality in experiences of physics education: Secondary school girls' and boys' perceptions of their physics education and intentions to continue with physics after the age of 16. *International Journal of Science Education*, *35*(11), 1824–1845. https://doi.org/10.1080/09500693.2012.762699

Müller, A., & Brown, A. (2022). An evidence-based approach to tasks in science education. Meta-analytic and other quantitative results. *Progress in Science Education*, *5*(1), 6–32. https://doi.org/10.25321/prise.2022.1275

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology*, *25*(1), 3–53. https://doi.org/10.1006/ceps.1999.1019

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* (pp. 115–143). American Psychological Association.

Nagy, G., Watt, H. M. G., 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), 482–506. https://doi.org/10.1111/j.1532-7795.2010.00644.x

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. National Academies Press.

National Research Council. (2015). *Identifying and supporting productive STEM programs in out-of-school settings*. National Academies Press.

Naylor, F. D. (1981). A state-trait curiosity inventory. *Australian Psychologist*, *16*(2), 172–183. https://doi.org/10.1080/00050068108255893

Nissen, J. M., & Shemwell, J. T. (2016). Gender, experience, and self-efficacy in introductory physics. Physical review. *Physics Education Research*, *12*(2), 020105. https://doi.org/10.1103/PhysRevPhysEducRes.12.020105

O'Keefe, P. A., Horberg, E. J., & Plante, I. (2017). The multifaceted role of interest in motivation and engagement. In P. A. O'Keefe & J. M. Harackiewicz (Eds.), *The science of interest* (pp. 49–67). Springer. https://doi.org/10.1007/978-3-319-55509-6_3

OECD. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA*. OECD Publishing.

OECD. (2007). *PISA 2006-Volume 2: Data*. OECD Publishing.

OECD. (2015). *The ABC of gender equality in education: Aptitude, behaviour, and confidence*. OECD Publishing. https://doi.org/10.1787/9789264229945-en

OECD Global Science Forum. (2006). *Evolution of student interest in science and technology studies [policy report]*. OECD Publishing.

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), 181–206. https://doi.org/10.1207/s15326985ep4103_4

Oppenheimer, F. (1975). The Exploratorium and other ways of teaching physics. *Physics Today*, *28*(9), 9–13. https://doi.org/10.1063/1.3069149

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), 1097–1119. https://doi.org/10.1002/tea.3660311005

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), 161–171. https://doi.org/10.1002/(SICI)1098-237X(199704)81:2<161::AID-SCE3>3.0.CO;2-D

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), 1049–1079. https://doi.org/10.1080/0950069032000032199

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), 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Palmer, D. H. (2009). Student interest generated during an inquiry skills lesson. *Journal of Research in Science Teaching*, *46*(2), 147–165. https://doi.org/10.1002/tea.20263

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. 224–246). Cambridge University Press.

Pawek, C. (2009). Schülerlabore als interessefördernde außerschulische Lernumgebungen für Schülerinnen und Schüler aus der Mittel- und Oberstufe [Science outreach labs as interest-fostering learning environments for secondary school students] [Doctoral dissertation, Christian-Albrechts Universität Kiel]. URN: urn:nbn:de:gbv:8-diss-36693

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 [the situational epistemic interest in physics topics of girls and boys after visiting a science outreach lab]. *Zeitschrift für Didaktik der Naturwissenschaften*, *24*(1), 279–285. https://doi.org/10.1007/s40573-018-0073-z

Ramsey, L. R., Betz, D. E., & Sekaquaptewa, D. (2013). The effects of an academic environment intervention on science identification among women in STEM. *Social Psychology of Education*, *16*(3), 377–397. https://doi.org/10.1007/s11218-013-9218-6

Randler, C., Ilg, A., & Kern, J. (2005). Cognitive and emotional evaluation of an amphibian conservation program for elementary school students. *The Journal of Environmental Education*, *37*(1), 43–52. https://doi.org/10.3200/JOEE.37.1.43-52

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1, 2nd ed.). Sage Publications.

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. Retrieved from. https://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf

Rodenhauser, A. (2018). Bilinguale biologische Schülerlaborkurse. Konzeption und Durchführung sowie Evaluation der kognitiven und affektiven Wirksamkeit [Bilingual biology outreach lab courses. Conception and implementation as well as evaluation of their cognitive and affective effectiveness] [Doctoral dissertation, Bergische Universität Wuppertal]. URN: urn:nbn:de:hbz:468-20161108-105143-7

Rodenhauser, A., & Preisfeld, A. (2018). Intrinsic motivation in bilingual courses on molecular biology and bionics in an out-of-school laboratory. In N. Gericke & M. Grace (Eds.), *Challenges in biology education research. A selection of papers presented at the XIth conference of European researchers in didactics of biology (ERIDOB).* University Printing Office.

Röllke, K., Maak, A. L., Wenzel, A., & Grotjohann, N. (2020). Experimental workshops in student labs and at school—What influence do location and personality traits have on Learners' motivational variables? *Pedagogical Research*, *5*(1), 1–9. https://doi.org/10.29333/pr/6338

Rotgans, J. I., & Schmidt, H. G. (2017). Interest development: Arousing situational interest affects the growth trajectory of individual interest. *Contemporary Educational Psychology*, *49*, 175–184.

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), 530–540.

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), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Scharfenberg, F. J., & Bogner, F. X. (2010). Instructional efficiency of changing cognitive load in an out-of-school laboratory. *International Journal of Science Education*, *32*(6), 829–844. https://doi.org/10.1080/09500690902948862

Schiefele, H. (1978). *Lernmotivation und Motivlernen: Grundzüge einer erziehungswissenschaftlichen Motivationslehre [learning motivation and motive learning: Basics of an educational motivation theory]*. Ehrenwirth.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review*, *13*(1), 23–52. https://doi.org/10.1023/A:1009004801455

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* (2nd ed., pp. 34–54). Routledge.

Singh, M. (2009). *Why recognition matters–valuing informal and non-formal learning across the north–south divide*. Hamburg UNESCO Institute of Lifelong Learning.

Sjøberg, S., & Schreiner, C. (2010). The ROSE project: An overview and key findings [final report]. Retrieved from https://roseproject.no/?p=63.

Skydsgaard, M. A., Møller Andersen, H., & King, H. (2016). Designing museum exhibits that facilitate visitor reflection and discussion. *Museum Management and Curatorship*, *31*(1), 48–68. https://doi.org/10.1080/09647775.2015.1117237

Stocklmayer, S. M., Rennie, L. J., & Gilbert, J. K. (2010). The roles of the formal and informal sectors in the provision of effective science education. *Studies in Science Education*, *46*(1), 1–44. https://doi.org/10.1080/03057260903562284

Streller, M. (2015). The educational effects of pre and post-work in out-of-school laboratories [Doctoral dissertation, Technische Universität Dresden]. URN: urn:nbn:de:bsz:14-qucosa-192707

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, *49*(4), 515–537. https://doi.org/10.1002/tea.21010

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Taconis, R., & Kessels, U. (2009). How choosing science depends on Students' individual fit to 'science culture'. *International Journal of Science Education*, *31*(8), 1115–1132. https://doi.org/10.1080/09500690802050876

Thomas, C. L. (2012). Assessing high school student learning on science outreach lab activities. *Journal of Chemical Education*, *89*(10), 1259–1263. https://doi.org/10.1021/ed200320g

Todt, E., & Schreiber, S. (1998). Development of interests. In L. Hoffmann, A. Krapp, K. A. Renninger, & J. Baumert (Eds.), *Proceedings of the Seeon conference on interest and gender*. IPN.

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*, *51*(4), 479–484. https://doi.org/10.1111/1467-9884.00332

UNESCO. (2017). *Cracking the code: Girls' and women's education in science, technology, engineering, and mathematics (STEM) [Final report]*. United Nations Educational, Scientific and Cultural Organization.

Usher, E. L., Ford, C. J., Li, C. R., & Weidner, B. L. (2019). Sources of math and science self-efficacy in rural Appalachia: A convergent mixed methods study. *Contemporary Educational Psychology*, *57*, 32–53. https://doi.org/10.1016/j.cedpsych.2018.10.003

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), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

Van Winkle, C. M. (2012). The effect of tour type on visitors' perceived cognitive load and learning. *Journal of Interpretation Research*, *17*(1), 45–57. https://doi.org/10.1177/109258721201700104

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), 574–588. https://doi.org/10.1177/1745691611421204

Walberg, H. J., & Tsai, S.-L. (1983). Matthew effects in education. *American Educational Research Journal*, *20*(3), 359–373. https://doi.org/10.3102/00028312020003359

Wang, C., & Frye, M. (2019, March 16). miniGEMS 2018 Summer Camp Evaluation: Empowering Middle School Girls in STEAM. Paper presented at 2019 IEEE Integrated STEM Education Conference (ISEC), Piscataway, NJ. DOI: https://doi.org/10.1109/ISECon.2019.8881981

Werquin, P. (2010). *Recognising non-formal and informal learning outcomes, policies and practices: Outcomes, policies and practices, 2009*. OECD Publishing.

Wherry, R. J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Annals of Mathematical Statistics*, *2*(4), 440–457. https://doi.org/10.1214/aoms/1177732951

White, R., & Gunstone, R. (1992). *Probing understanding*. Routledge. https://doi.org/10.4324/9780203761342

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review*, *12*(3), 265–310. https://doi.org/10.1016/0273-2297(92)90011-P

Woithe, J. (2020). Designing, measuring and modelling the impact of the hands-on particle physics learning laboratory S'Cool LAB at CERN [Doctoral dissertation, Technische Universität Kaiserslautern]. https://doi.org/10.13140/RG.2.2.13201.66403

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 [Doctoral dissertation, University of Michigan].

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*, *45*(9), 1036–1058. https://doi.org/10.1002/tea.20195

How to cite this article: Woithe, J., Müller, A., Schmeling, S., & Kuhn, J. (2022). Motivational outcomes of the science outreach lab S'Cool LAB at CERN: A multilevel analysis. *Journal of Research in Science Teaching*, *59*(6), 930–968. https://doi.org/10.1002/tea.21748