Effectiveness of science outreach labs with and without connection to classroom learning: Affective and cognitive outcomes

Alexander Molz,[1,*] Jochen Kuhn,[2] and Andreas Müller[3]

[1]*Department of Physics, University of Kaiserslautern, Erwin-Schrödinger Str. 46, 67663 Kaiserslautern, Germany*

[2]*Faculty of Physics/Chair of Physics Education, Ludwig-Maximilians-University Munich, Theresienstr. 37, 80333 Munich, Germany*

[3]*Faculty of Sciences/Physics Department and Institute of Teacher Education, University of Geneva (University Institute for Teacher Education), Boulevard du Point d'Arve 40, CH - 1205 Geneva, Switzerland*

(Received 15 July 2022; accepted 27 September 2022; published 13 December 2022)

Affective and cognitive outcomes of science teaching in schools may be enhanced by science outreach labs (SOLs). Particularly interesting is whether and how to connect SOL visits to in-school science learning. Recent studies—among others, in this journal—have confirmed SOLs' positive affective effects. However, research remains inconclusive regarding cognitive outcomes, the effects of connecting SOLs to classroom teaching, and effects beyond the short term. This study more closely investigates the short- and medium-term effectiveness of SOLs on affective (interest, self-concept) *and* cognitive variables (conceptual and procedural knowledge), as well as the effects of integrating lab work with school-based preparation and post-lab activities. A quasi-experimental intervention study via a repeated measures design and two SOL treatment groups (TG1/2: with and without integration), and an in-school control group (CG; also with integration of pre- and postlab activities, in accordance with good practice) was undertaken. It took place in a lab work unit on "pressure and buoyancy" for the lower secondary level, with identical instruction and the same instructor across groups, and with several further control measures. The main findings are as follows: (i) SOLs can have substantial learning and affective outcomes [pre-post learning gains: Cohen's $d > 1$; increases in interest ($d > 0.8$) and self-concept ($d \approx 0.5$) immediately after the visit]. (ii) Learning gains were still present at medium term ($d > 0.7$) but not affective ones. (iii) Integration with classroom teaching was necessary for learning (TG1 vs TG2: $d = 0.72$), and not harmful for interest. (iv) Learning and affective outcomes of an integrated SOL are as good as those of a well-prepared classroom setting, but not better. (v) No interactions of the outcome variables with gender and other covariates were found; the lab work units appeared suitable for both sexes and different kinds of learners. These findings are discussed with respect to current theories of learning and interest development, and SOLs' added value to science education.

DOI: 10.1103/PhysRevPhysEducRes.18.020144

I. INTRODUCTION: FROM ENJOYMENT TO LEARNING?

An enjoyable and successful visit experience is an important outcome because it can predispose the learner to engage in further cognitive learning [1].

Over the last decades, out-of-school learning opportunities (OSLO[1]) offered by science centers, science museums, and various research institutions have experienced a very strong development and are increasingly recognized as an integral part of science, technology, engineering, mathematics (STEM) education throughout the world [2–5]. Friedman [6] described the evolution of science centers and museums toward the hands-on, interactive format they

*Corresponding author.
amolz@physik.uni-kl.de

[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; in fact, such a connection appears as an essential success factor for OSLOs (Sec. II. B) and is the research focus of several recent studies, including the present one.

provide nowadays in most cases, and which is considered particularly promising for audiences of high school learners [7,8].

A specific form of these opportunities, provided by a wide range of host institutions from science departments of universities to science centers, and in various formats (permanent, temporary, mobile, etc.) are "science outreach labs" (SOL) [9]. They are based on experimental hands-on activities and active lab work by the participants, most often within workshops lasting a few hours to a full day [10–14] and sometimes connected to other educational offers (lectures, lab visits, etc.; del Barco [15]; APS [16]).

A main purpose of these initiatives is to promote positive affective and attitudinal effects, such as an enjoyable science experience, curiosity, interest, and openness for specific topics, or science in general [7,8]. Affective and attitudinal aspects are also considered as an important component of scientific literacy (attitudes of future citizens, interest in learning about science, etc. [17]). Recently, the role of out-of-school physics experiences to foster participation and the development of a "physics identity" of underrepresented groups in society was highlighted [18,19]. An enjoyable science experience can provide a starting point for further learning (see the quote by Rennie [1] at the beginning of this section). Enjoyment is also an essential factor of an out-of-school learning offer from the point of view of teachers. In a series of studies with hundreds of teachers over several years, enjoyment was constantly indicated by roughly 80%–90% of them as one of the major benefits for pupils [20].

However, a major challenge for out-of-school learning opportunities is providing evidence that they have positive effects (i) on the affective *and* cognitive level and (ii) beyond the short term. For instance, concerning field trips, Rop [21] stated "that there is a great need for studies that concentrate on cognitive learning outcomes." More recently, De Witt and Storcksdieck [22], in a review on excursions (in the sense of school trips in general), concluded "that major and measurable cognitive gains tied to scientific concepts are difficult to achieve during the short time span of most excursions." As for long-term effects, Falk and Heimlich [23] stated that most research "has been focused on the short-lived motivations engendered by the learning environment."

Of particular interest is the question of whether and how out-of-school learning places should be connected to formal science learning taking place in schools [8,24]. Indeed, in view of around 15000 hours a student has spent at school by the completion of secondary school [25], it is not obvious how a disconnected intervention of a few hours spent in an OSLO, that is, of the order $\sim 10^{-4}$ in duration, should have a major impact on motivation or attitudes. Moreover, there is solid research evidence that both motivation and learning need time and consistent, connected experiences to develop [26,27], see Sec. II. B for further development of this. The purpose of the present work is to contribute to the understanding of the effects and the effectiveness of out-of-school science learning offered by the investigation of a specific setting of an SOL connected to regular classroom teaching at school, using measures for both affective and cognitive outcomes, and following their development beyond the short term.

II. RESEARCH BACKGROUND

A. Out-of-school science learning: Affective and cognitive effects

1. General rationale and affective outcomes

On a more general level, OSLOs can help to foster ongoing, lifelong attitudes and learning processes in the sense of scientific competence [28,29], especially when coordinated with in-school (formal) learning [8,24,30]. Braund and Reiss [7] discussed the following ways in which out-of-classroom contexts can foster science learning (their formulations are given in italics):

"Extended and authentic practical work" and *"Access to rare material"*: Most science out-of-school learning opportunities provide learning experiences strongly based on practical and experimental activities, known to be a good way to foster interest and achievement in science [31]. Often, this is referred to as a "hands-on" character of OSLOs [24,32], recently reviewed for its research background in detail by Besse *et al.* [33]. Braund and Reiss [7] also mentioned "*access to big science*", an aspect provided, for example, by the SOLs at CERN [11], the Paul-Scherrer-Institute [34], or Fermilab [13].

"Attitudes to school science, stimulating further learning": This is consistent with the research cited in Sec. I and one of the positive OSLO effects most consistently stated in the field (see also the quote by Rennie [1] at the beginning of Secs. I and II. B. 4).

"Improved development and integration of concepts": This represents an important aspect for the present study to which we will return below (Sec. II. B. 2).

"*Collaborative work*": In many cases, "collaborative work" is part of the learning setting at SOLs, in accord with existing evidence for regular science classrooms ($d = 0.95$ for science learning with vs without cooperation for primary and secondary school level [35]; see also Ref. [36]).

Solid evidence supports various positive affective outcomes of OSLOs ([28,37]; see also Sec. II. B. 4). Empirical data from a large variety of OSLOs in several European countries provide strong evidence that they are, indeed, able to create a high level of enjoyment or general appreciation (assessed in various ways at different sites, for example, by Likert scales, single items, a "grade" for the visit, etc.). As Fig. 1 shows a range of 70%–90% of the

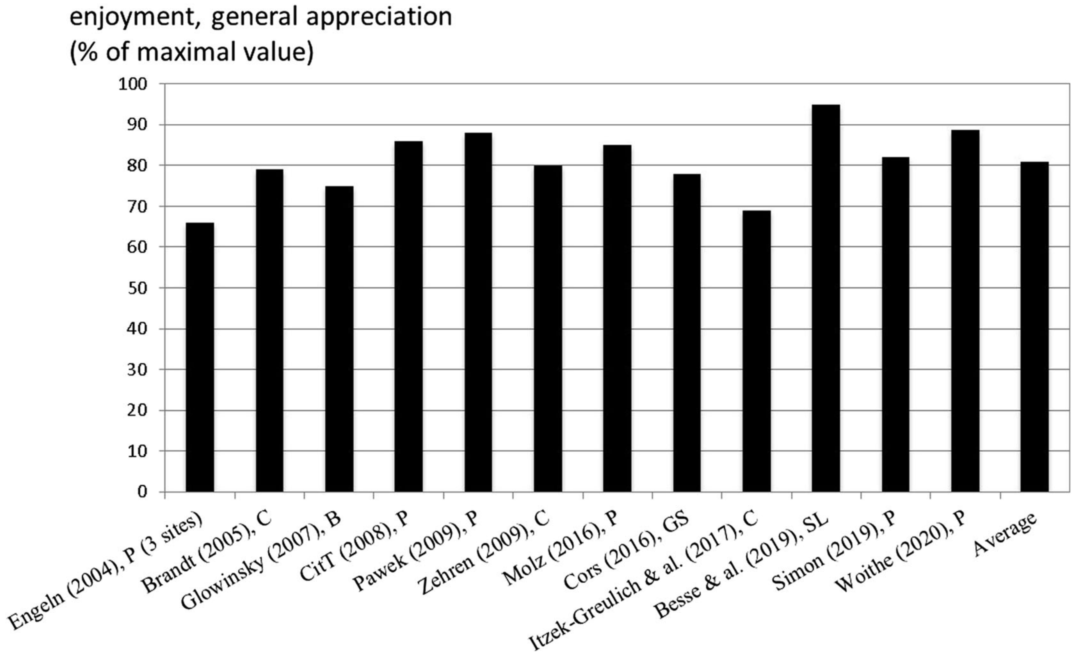

FIG. 1. Enjoyment or general appreciation of various science outreach labs as percentage of maximal possible score value ("POMP" [38]). Abbreviations: B: biology; C: chemistry; P: physics; GS: science (general); SL: scientific literacy; rightmost column: average [32,33,39–48].

maximal possible value of the measure at each site is consistently obtained, with an average around 80%. Note that these results hold across several countries (e.g., France, Germany, Switzerland), target groups (primary school to general public), various settings (single and multiple visits, degree of guidance) and disciplines (biology, chemistry, physics), and across more than a decade. Recently, Itzek-Greulich and Vollmer [32] provided strong evidence for positive affective outcomes in a large sample study. This clearly shows the high potential of out-of-school learning opportunities to provide an enjoyable science experience.

Moreover, SOLs have been shown to foster the science interest of less interested students: The science interest gap between initially highly and weakly interested students (median split) decreases from $d = -2.33$ before to $d = -0.49$ after a SOL visit [43]. Similarly the well-known "gender gap" in science interest between girls and boys was shown to decrease from d (long term) $= -0.85$ to d (after visit) $= -0.51$ [43]. A similar effect was found for the self-concept of girls [d (long term) $= -0.9$, d (after visit) $= -0.7$; [49]]. Note that while the effects on the gender gap may not appear large, they refer to rather stable traits, if not so say deeply entrenched attitudes [50], and the results thus show that even a short intervention in a SOL can initiate a change of them.

Taken together, the above arguments and findings suggest that out-of-school learning places can foster affective outcomes and attitudes in a way traditional science teaching at school cannot; we call this the "place" hypothesis.

2. Cognitive outcomes

Classical formats of out-of-school learning, where cognitive outcomes have been studied to a certain extent, are outdoor field trips in the life and earth sciences and planetarium visits in astronomy.

Outdoor field trips and excursions[2] are well-established elements of the life and earth sciences, as experience "in the field" is constitutive both for these disciplines themselves and for their teaching [51]. Central features are authentic experience, inquiry, and experiential learning [52,53], very much in common with SOLs.

Prather [54], in a review on early research on field trips, concludes that it "clearly supported the use of field instruction for both factual and conceptual learning as well as for affective objectives" and that "compared to other traditional teaching techniques, field trips may provide an especially rich stimulus setting for content learning." However, no quantitative results are included in this review. Fuller *et al.* [51], in an international perspective on fieldwork in geography, agree that is an effective means of learning, but add "just how effective it is compared with other methods of learning remains in need of investigation." In the following discussion, the aspect of "effectiveness" is completed by specification of effect sizes wherever possible; that is, we include values of Cohen's d either for

[2]There is also a broader understanding of field trips including e.g., museum visits which we do not consider here.

comparisons of different treatment groups or for pre-post comparisons (the latter are denoted as d^* for clarity).[3]

Foundational work on cognitive effects of field trips in the life and environmental sciences was done in the US by Falk and co-workers. They found large pre- and posteffects ($d^* \approx 0.8 - 1.7$) for various cognitive outcomes (factual and conceptual knowledge) [58–60]. In these studies by Falk *et al.* various visitor groups and types of field trips were compared, to which we come back in Sec. II. B. 3. Another study in the US provided evidence for strong effects on the learning of ecological concepts ($d^* > 1.5$; [61]), and that they were still present one month after the intervention. Killermann reported comparable effect sizes in Germany ($d^* = 1.55$; [62]). With one exception (high school; [61]), all these studies were about young learners (10–13 yr).

For field trips in the Earth sciences, the study of Orion and Hofstein [63] showed considerable effects for higher order learning outcomes (problem solving related to field trip phenomena), while those for more factual knowledge were smaller ($d^* = 0.56$, $d^* \approx 0.3$, respectively; 9–11 grade). For planetarium visits, a meta-analysis of 19 studies showed a small but significant average effect ($d = 0.28$) across all age groups and for a wide range of learning objectives [64]. DeWitt and Storksdieck [22] concluded that field trips "can certainly lead to cognitive outcomes" but that "these gains are often quite small and dependent on the degree to which students were readied to engage with content." Other conditions of successful learning in out-of-school settings are discussed in the literature, to which we will come back below.

We now turn to studies on cognitive outcomes specifically of SOLs, for which only a few authors have provided quantitative measures. Zehren [44] investigated the impact of a longer-lasting intervention over 2 years with repeated visits to a chemistry SOL (7 to 9 visits, 2 h of experimentation each) on a range of quality criteria of student experiments. Comparing to a control group without SOL visits (however experimental lessons at school), he obtained large to very large effect sizes (among others, level of experiment-related questions: $d = 1.32$; planning and execution: $d = 1.2$; analysis and interpretation: $d = 0.83$; all obtained by expert rating in a final test session about a new experiment).

[3]Reported values of Cohen d are either directly published ones, or inferred from other published statistics according to standard procedures [55]. Usual effect-size levels (as established from comparison of a great many of studies in different areas) are small ($0.2 < d < 0.5$), medium ($0.5 \leq d < 0.8$) or large ($0.8 \leq d$) [56]. Recently, Hattie [57] has introduced the "hinge point" of $d = 0.4$ as comparison value (it is the average of all effect sizes reported in his meta-meta-analysis). For correlations, the corresponding levels are $0.1 < r < 0.3$ for weak, $0.3 < r < 0.5$ moderate and $r \geq 0.5$ for strong effects, respectively.

Damerau [65], in a study on a SOL about cell and molecular biology, found very large pre-post effects for learning about different topics (unicellular organisms: $d^* = 2.0$; anaerobic cell metabolism: $d^* = 1.59$; genetic fingerprinting: $d^* = 1.52$). Hirth [66] investigated an SOL offer about sound and distinguished in his learning test understanding of (i) concepts and (ii) graphical representation, as well as (iii) explanations as important components of scientific understanding, and found large effect sizes for all ($d^* = 1.23$, 1.08, 1.1, respectively). As for learning beyond the short term, the studies of Damerau [65] and Hirth [66] also provided evidence for an at least medium-term stability of the learning effects by SOL courses ([65]: $d^* = 0.7$, 1.0 for two biology topics, $\approx 2.5 - 3$ months after the intervention; [66]: $d^* = 0.9 - 1.0$ for various learning components regarding the topic of sound, 1–2 months after the intervention). Recent work by Itzek-Greulich and coauthors [32,67] will be discussed in Sec. II. B. 4. In view of the above, it can be concluded that out-of-school learning offers in general, and SOLs in particular, can also foster cognitive outcomes for science learning. In view of general arguments put forward in the literature ("extended practical work," "improved development of concepts," see above), it can be hypothesized that they can do this better than traditional science teaching at school; thus, there is a "place" hypothesis regarding cognitive outcomes. Note that while actual empirical evidence concerning the comparison of cognitive outcomes to learning at school is scarce, this place hypothesis is an implicit hypothesis of many operating institutions of SOLs, believing that their investment leads to a substantial added value beyond science learning at school.

B. Connecting out-of-school and in-school learning

1. General arguments and teacher's views

One of the most often stated features for the quality and success of out-of-school learning offers, for both affective and cognitive outcomes, is that they should be connected to in-school learning [8,24,30]. This holds true for all formats of OSLO discussed here, field trips [52], planetarium or astronomy outreach [68] and SOLs [44,69]. Two major forms of such connections are (i) pre- and postvisit activities at school ("integrated approaches," [70]) and (ii) links of the OSLO content and activities to the curriculum ("curricular links"), where the former most often implies the latter.

Teachers perceive these connections as a decisive factor of success. According to a report by the US National Academies of Sciences' Committee on Science Learning "[…] teachers' perceptions about curriculum fit and pre- and postvisit activities emerge as the most frequently cited factors to impact the development of disciplinary specific knowledge from field trips" [30]. Indeed, in the study by Anderson and Zhang [71], curriculum fit was found to be the most important planning factor from the teachers' point of view. In several other studies, 80%–100% of teachers also stated the importance of curriculum links [72] and of pre- and postactivities [73,74].

Rennie [28] stated that early advice intended to help teachers organize effective field trips given by museum educators "[…] emphasized that planning before the trip and "recapitulation" afterward were almost as important as the trip itself. Research has confirmed this advice." Consistent with these statements about good practice by educators in and out of school, we will now turn to three strands of research providing theoretical background and evidence for the affective and cognitive benefits of connecting out-of- and in-school learning.

It is also noteworthy that for science teaching at schools an "integrated setting" of student's experiments with preparation and follow-up work is to be considered as a common standard, based on available research evidence and good practice [75–78]. An isolated lab work would not be considered a recommendable educational approach.

Motivation and willingness to engage in further instruction are most likely to be the important affective outcomes of a visit. [1]	
Interviewer:	*What did you learn about these things—in school, here?*
Student:	*[A]bout the Kennedy and Sputnik things, we learned that last year in science. We spent months on that, and I am shocked that I remembered it. And it was pretty cool. I'm not really good at science, so that was really like a big thing for me.*
(Interview study at the "Smithsonian", [87])	

FIG. 2. Connecting learning in and out of school: high potential of OSLOs as "catch" events (top) and for positive affective impact of "experience of competence" (bottom).

2. Development of interest and knowledge as sustained processes

Person-Object approach to interest and interest development.—The person-object theory of interest and interest development (POI) is "explicitly oriented to the demands of educational practice" [79]. It is widely and successfully used in science education [80] and is also well suitable to describe the impact of OSLOs on interest [81]. Its key ideas are as follows [79,82]:

(i) Interest is understood as an *interaction* between a person and an object of interest. This may sound self-evident, but it implies that intrinsic (person-related) and extrinsic (situation or object-related) factors are taken into consideration in a single theory.

(ii) Interest is also understood in its *development* from a person's current state to a lasting trait. This may again appear obvious, but it means that the temporal development in interest can be captured in a differentiated way, in a unified theory. Development of interest can be tracked starting from a state of "current" interest aroused by the features of a given situation, its consolidation ("catch and hold"), through eventually formation of a person's "dispositional" or long-term interest.

The theory has been repeatedly discussed and used in relation to science education [79,80,82] in general, and to SOLs in particular [32,43,69,83,84]. By their very definition, SOLs are meant to engage people in interesting experiences and activities (see Sec. II. A. 1), and thus to spark situational interest. POI then predicts that dispositional interest can develop out of this triggering situation, if integrated in appropriate catch-hold processes. One way to do this are post-visit activities in the classroom, related to and developing the content and experiences met at the OSLO [85,86]. The high potential of out-of-school learning offered as "catch" events belongs to their very rationale ([1]; see quote in Fig. 2). Note, however, that the flow of events in the catch-hold process can also be from in- to out-of-school learning. A quote from an interview study by the "Smithsonian Institution" [87] on "enhancing the visits" of school groups nicely illustrates how an opportunity of "connected" learning in an out-of-school setting can provide a memorable, lasting experience of satisfaction and encouragement for a pupil ("experience of competence," see above).

Prior knowledge and cumulative learning.—Prior knowledge in a domain is a strong predictor of learning in general ($d = 0.67$), and of science in particular ($d = 0.8$, [57]). In a wider perspective, the notions of "cumulative learning" or "knowledge integration" take account of the temporal development of knowledge as a fundamental dimension of learning [27]. They are understood as the "sequential development of knowledge and skills" and emphasize the role of knowledge structures as systematic, well-structured bodies of knowledge [27].

Specifically, several studies have provided clear evidence that opportunities for knowledge integration provided by adequate preparation and follow-up work are also essential for the effectiveness of science learning in laboratory lessons ([88,89]; see Augustian and Seery [90] for a recent review), or otherwise, that without such a framework, chances are high that it stays merely "a set of disconnected actions to be followed" [91].

The importance of prior knowledge and cumulative learning was also repeatedly stressed for out-of-school learning settings [3,84]. Rennie *et al.* [92] state that "[l]earning is a cumulative process involving connections and reinforcement between the variety of learning experiences a person encounters in his life" (see also Sec. II. B. 4).

3. Novelty

Early research about novelty and its potentially distracting effects provides further evidence for the educational advantages of connecting out-of- and in-school learning experiences. In the foundational work of Falk and coworkers already mentioned in Sec. II. A. 2, they investigated the influence of novelty, or unfamiliarity, on the learning effects of field trips. In several studies, they compared a "familiar" group of pupils who had already visited a nature center, but for another topic, to an "unfamiliar" group who visited the center for the first time. They found considerably better learning for the familiar group [58–60] and concluded that for the unfamiliar group, "exploration and setting-oriented learning took precedence over conceptual learning."

Based on these findings, the idea of novelty-reducing measures appeared in the literature. Kubota and Olstadt [93] found a positive effect of vicarious exposure preparation (an audio-visual presentation showing children in typical situations of a visit and providing orienting comments).

Anderson and Lucas [94] studied the effect of previous visits and of an orienting preparation (about location and organization, but not about details about individual exhibits) on learning about various physics topics in a science center. Both previous experience and orienting preparation were shown to foster learning ($d = 0.7$ and $d = 0.6$, respectively). They summarize that while these are medium effect sizes, "their educational significance lies in recognizing how simply they were realized" (either previous experience or a brief pre-visit orientation).

It should be mentioned that beyond treating novelty as a single, one-dimensional factor, as in the above work by Falk and others, Orion and co-workers have extended it to a "novelty space" with several dimensions [63,95]. The idea of novelty space is closely linked to that of adequate preparation, "which will reduce the novelty space to a minimum and thus facilitate meaningful learning during the field trip" [95]. The differentiation according to these dimensions will not be investigated in the present research; the interested reader is referred to Ref. [95] for a study on this aspect.

The results discussed above highlight the positive effects of providing orientation and preparation for an OSLO visit. Of course, one could believe that a structured and well-prepared learning environment resembles too much a usual classroom or school atmosphere and thus is detrimental to a fundamental objective of OSLOs, viz. arising interest and a spirit of discovery. It is one purpose of the present paper to study whether the structured approach of an integrated SOL and school setting is in fact detrimental to affective factors or not.

4. Recent empirical results for connected settings of science outreach labs

The research presented in Sec. II. B. 2 provides theoretical foundations and empirical evidence that the development of knowledge as much as that of interest are long-term processes that can be (much) better fostered by single, short events like an OSLO visit when these are integrated in a systematic, continuous setting of related preparatory and/or post-visit activities ("connected settings"). Section II. B. 3 provides theory and evidence that an OSLO visit has to be safeguarded against distraction and cognitive overload to fully develop its educational potential, again emphasizing the helpfulness of preparation. On the basis of this previous research, a series of intervention studies specifically about various forms and aspects of connected settings in SOLs was carried out in the last decade, to which we turn know.

Glowinski [41,69], in an evaluation study on SOLs for molecular biology, provided evidence for the positive effects of preparation on several affective variables, in particular several components of general interest in the SOL offer, and of specific interest related to its topic, with effect sizes mostly in the range of $0.4 \lesssim d \lesssim 0.7$. Most pronounced was the effect of preparation on interest for experimentation in general ($d = 0.68$). Additionally, the effect of post-visit work on some of these affective variables was also studied and turned out to be positive (e.g., interest in the specific experiments done at the SOL, $d = 0.58$ or the topic treated there, $d = 0.68$). Pawek [43] found no direct effects of this kind, but nevertheless that the better learners felt prepared, the more they felt competent for the requirements of the SOL visit ($d = 0.46$). Note that all preceding results are correlational, that is, studying the association between the perception by learners and the variable of interest. Very few studies in this area had an interventional design, that is, actually compared a connected setting of an out-of-school learning offer to a nonconnected one.

Huwer [96], in another investigation in the same (bio) chemistry SOL as in Ref. [41], reported large to very large effect sizes for various topics when compared to a control group without teaching on these topics (fats and oils: $d = 2.04$, water cleaning treatment: 1.87, sugar and sweeteners: 3.8). Streller [97] investigated the impact of an online preparation for an SOL unit in physics (magnetism) and observed small to medium size effects on several affective variables (most pronounced: emotional component of situational interest: $d = 0.80$) and (perceived) comprehensibility ($d = 0.49$).

TABLE I. Research hypotheses.[a]

Hypothesis 1	An integrated setting of a science outreach lab yields *greater effects on students' learning (conceptual and procedural)* than (a) a setting without integration. (b) a lab work unit at school
Hypothesis 2	An integrated setting of a science outreach lab *yields greater effects on students' interest and self-concept* than (a) a setting without integration (b) a lab work unit at school.
Hypothesis 3	The positive effects of an integrated setting of a science outreach lab (a) on learning (b) on interest and self-concept Last beyond the short term and can still be detected several weeks after the end of the intervention.

[a]Note that while an advantage over labwork in schools is part of the rationale and a kind of current working hypothesis of SOLs, it is less justified by extant research than the advantage of integration; the hypotheses Hyp1b and Hyp2b, thus, could also have been formulated as more open research questions. For simplicity, we decided not to do so.

A very detailed study of this kind was undertaken by Itzek-Greulich and co-authors, who responded in an impressive way to the exhortation (see above) for better methodological quality in the field (use of five learning measures, cluster randomization of a large sample ($N \gtrsim 400$ in each of four comparison groups), multilevel analysis, etc. [32,67]). To the best of our knowledge, it is hitherto the only well-controlled intervention study on the effects of an integrated setting of an SOL unit (organic chemistry of carbohydrates) as compared to learning the same content in an SOL without integration, and in school. The integrated setting was composed of a theoretical introduction at school, the experimental part at the SOL, and a summary again at school. Note, however, that the lesson script in the SOL condition was developed by its staff, while that in the "school" condition it was developed by a teacher (providing additional elements typical for experimental sessions at school, for example, introductions to experiments and a workbook structure). The integrated setting had a comparable but shortened script to meet time limitations during the school part. Thus, while learning content and experiments were identical in this study [32,67], lessons scripts were not, reducing comparability. All three learning settings were compared to a control group not taught on the topic in question. For the integrated setting, they obtained an effect size for specific content knowledge of $d = 0.72$ (comparison with control group without instruction), and for specific experimental knowledge (a test of procedural knowledge about the techniques seen in the lab work) of $d = 0.37$ [67]. For the other learning measures, the difference turned out to be small or not significant. Moreover, comparing the integrated setting to a "school only" setting, learning in the latter turned out to be better for all four measures.

The test of specific content knowledge was a fill-in-the-blank test, which is considered to capture lower-order learning [98,99], and it did not refer to existing research about conceptual understanding of the subject matter, in particular not addressing known pre- and misconceptions [100,101]. For pre-post comparisons, the effect sizes for factual and experimental knowledge are very small ($d^* \lesssim 0.1$, as inferred from [67], Table 1)[4]; moreover, the overall effect for specific content knowledge was reported to be even negative[5] across learning groups in ([102], $d^* = -0.6$), while no pretest results and pre-post changes for this measure were reported in [67]. Across the life, environmental, and Earth sciences, and for astronomy, pre-post effect sizes of OSLOs for learning are, however, regularly around 0.8 and larger, often attaining the conventional threshold for a large effect (see Sec. II. A. 2). Despite the high-level methodology of the work by Itzek-Greulich *et al.* [67], the state of research on the possibility and the conditions of effective learning at SOLs thus still appears inconclusive for the measures used, the actual learning occurred, and the results found regarding the comparison of integrated, nonintegrated, and school only settings. This is a major reason for the present work.

C. Learner and setting characteristics

The dependent variables learning and motivation are influenced by several factors [103]. As stated in Sec. II. B. 2, prior knowledge is a strong predictor of learning in general, and of science learning at OSLOs in particular. This includes content-specific knowledge, as well as general background knowledge of the discipline in question.

[4]The apparent contradiction to the larger treatment-control group effect size at post-test is explained by the fact that the knowledge measures for the (non-intervention) control group at post-test are generally quite low, eg. for specific experimental knowledge considerably lower than at pre-test. Thus while the treatment group improved to a certain extent, the control group became worse to a similar or larger extent, producing an increase of Cohen d according to its definition.

[5]Note that these d values are negative as they refer to for interest *gaps*, e.g., between weakly (M_1) and highly interested students ($M_2 > M_1$, thus $M_1 - M_2 < 0$).

Another important predictor, especially of physics learning, is prior knowledge in mathematics [104]. Considerable correlations between mathematics and physics or science competences have been reported ($r = 0.48$ according to meta-analysis, [105], corresponding to $d = 1.1$). As both autonomous learning during lab work (experiment description, activity sheets) and testing are heavily based on language, reading language skills are also to be taken into account. Additionally, general (nonverbal) cognitive abilities have been found to be a strong predictor of learning and achievement ($d = 1.19$, [57]).

Moreover, the initial state of interest and self-concept as affective variables is also of importance; both are predictors of science learning (interest: $r = 0.3$, [106]; self-concept: $r = 0.43$, [57]; corresponding to $d = 0.63$ and 0.95, respectively), and their positive development belongs to the core objectives of out-of-school learning offers.

Gender differences regarding physics or physical science for both cognitive and affective variables have also been known for decades. For achievement, differences (in favor of boys or men) are small, but nonzero ($d = 0.32$, [107]). For interest and self-concept, a similar picture is obtained: (interest: $d = 0.4$, from the 5000 pupils UPMAP study in England, [108]; self-concept: $d = 0.27$, OECD average for science in general [109]).

Lastly, the important role of educators in a given learning setting has been repeatedly stressed. For learning at school [57] stated that "teachers are among the most powerful influences in learning." Similar emphasis has been given to the role of instructors in out-of-school learning settings ([110], Chap. 7). When comparing in-school and out-of-school settings, it is thus important to have the same person as educator to avoid confounding effects.

D. The present study

1. Purpose of the present study

As the preceding sections show, large positive affective and cognitive effects by out-of-school science learning offers are possible. Consistent with narrative, qualitative accounts of earlier years, several quantitative studies of the last few decades have found large effects for learning gains (pre-post comparison, $d^* \geq 0.8$), and in some cases, also in comparison to classroom teaching, in a wide range from small to very large effects (see Sec. II. A. 2, $0.3 \lesssim d \lesssim 1.3$). However, these findings and their analysis, in particular regarding learning, suffer from serious limitations: First, with the very small number of well-controlled studies and the absence of replication studies or meta-analyses (except for planetarium research), there is a need for further research on important specific features of out-of-school learning offers, such as the effects of connected settings. Second, there are limitations on the level of individual studies, in particular concerning design, instruments, lack of consideration of control variables, and underreporting (e.g., not providing effect sizes or sufficient data to compute them), as stated repeatedly by researchers in the field [67,111] and, for example, by the National Research Council (US), in a review on the state of affairs ("[i]mproving the quality of evidence on learning science in informal environments is a paramount challenge," [3]).

For research on connected and integrated settings of SOLs, which on the basis of the above arguments (Sec. II. B) is an approach of very high importance, Itzek-Greulich *et al.* [32,67] presented a high-quality comparison study but with small learning effects in either of the compared settings and reduced comparability between the settings due to different scripts (Sec. II. B. 4) and different educators in the comparison groups. Thus, there appears a research need for a well-controlled comparison study, with identical scripts for the SOL and lab work at school, and where sizable learning actually occurred. This is the main purpose of the present contribution. In view of the focus of SOLs on experimentation, and of its importance for science education in general [75,77], procedural understanding of the experiments and of the relations studied in them is one important aspect of learning studied in this work. Moreover, it is of interest to include the effects specifically on conceptual learning, using an appropriate measure, ensuring curricular validity (as one wants to compare to a school setting), and based on existing research. Kubota and Olstadt [93] have emphasized the need for such measures for learning at OSLOs as early as in 1991. Therefore, in this study, we include a research-based, curriculum-valid measure of learning with a focus on conceptual and procedural understanding.

Given that positive motivational outcomes[6] belong to the main purposes of out-of-school learning offers (Sec. II. A. 1; [3,7,8,32]) they are also investigated in this study. Lastly, if indeed positive cognitive and affective effects in a science outreach lab occur, it is of interest whether this holds beyond a short-term effect [49,114–116]. The present study thus includes a follow-up test one and a half months after the intervention. For both cognitive and affective effects, the study also includes several control measures and variables based on previous research (Sec. II. C).

2. Research hypotheses

In the present work, the impact of a science outreach lab on conceptual and procedural learning as cognitive outcomes and on interest and self-concept as affective outcomes was studied. In particular, we assessed the effects of an SOL with an integrated setting (preparation and follow-up work) as compared to an SOL without integration, and to a laboratory unit at school. Based on the research

[6]'Motivation' is often used as an umbrella term including interest and self-beliefs (especially in the context of education [112,113] and it is in this sense that we use the term here.

TABLE II. Sample composition (TG: Treatment group, CG: Control group).

	TG1 (SOL)	TG2 (SOL)	CG (School)	Total
Girls	34	28	20	82
Boys	29	24	18	71
Total	63	52	38	153

background developed above, the research hypotheses are presented in Table I.

III. MATERIALS AND METHODS

A. Study setting and sample

The study was conducted in the German state of Rhineland Palatinate in the years 2014–2016. Data were collected from 190 ninth-grade students (lower secondary level) in 8 classes at 3 academic-track schools in the German school system.[7] Complete class groups were randomly assigned to the individual groups. Because of incomplete datasets, data from 37 students had to be excluded, and we follow a complete case analysis (or "listwise deletion" [117]). This is considered as an acceptable approach, and, in particular, taking account of differences between participants with a potential effect on the outcomes and including them as covariates in the model (here initial understanding, interest and self-concept) is an effective way of reducing bias.[8] The resulting sample consisted of 153 participants ($M_{\mathrm{age}} = 14.6$ a, SD = 0.6 a; 46.4% boys). A sample breakdown is given in Table II.

Subject matter was "pressure and buoyancy," a standard topic according to the pertinent lower secondary school physics curriculum (see Sec. III. C for details). The teaching units took place either at the schools of the participant classes (control groups) or at the science outreach lab "iPhysicsLab"[9] of a technology-oriented university in Germany (treatment groups).

B. Study design

The intervention study had a quasi-experimental control-group design with repeated measures, see Fig. 3. A first treatment group (TG1) received the specific preparation lesson before and the specific follow-up lesson after the visit to the SOL. The second treatment group (TG2) also participated in the lab work session at the SOL but received

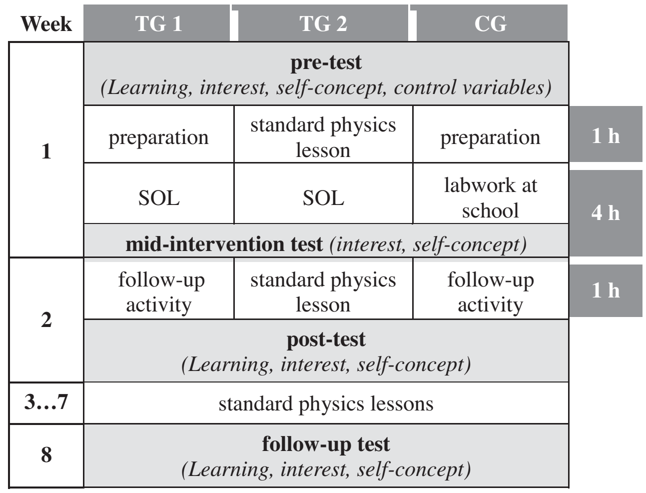

FIG. 3. Study design. The comparison between TG1 and TG2 is about a lab work unit with and without integration (Hyp1a, Hyp2a) and the comparison between TG1 and CG is about a lab work unit at a SOL and at school (Hyp1b, Hyp2b).

two standard physics lessons on the subject of pressure and buoyancy, one before and one after the visit. In contrast to the preparation and follow-up lessons, these two standard physics lessons were not specifically related to the lab work content. They were necessary to exclude possible time on task effects due to different intervention durations.

The control group (CG) carried out the experiments in school rather than at the SOL. Similar to the TG1, the CG received the preparation and follow-up lessons before and after the lab work unit. Note that an isolated lab work unit at school without preparation and follow-up work would not be considered a recommendable educational practice (see Sec. II. B. 1). Thus, the comparison should be carried out with respect to a school lab work unit with integration, not an artificially impoverished one without. Content, equipment, and the lesson plan of the lab work sessions for all three groups, as well as of the preparation and follow-up lessons for TG1 and CG, were identical.

The research hypotheses are related as follows to the study design: The comparison between TG1 and TG2 provides information on the effect of the specific integration of the SOL into standard physics lessons on the outcome variables (Hyp1a, Hyp2a; "integration hypothesis"). The comparison between TG1 and CG provides information about the effectiveness of the learning place itself (Hyp1b, Hyp2b). The delayed follow-up test provides information on the temporal stability of cognitive and motivational effects (Hyp3a, b).

A comment on the following question is appropriate here: Why study the effect of a SOL with experiments identical to school activities ("SOL = school" setting)?

First, there are in fact SOLs which such a setting. A strong reason for this is that while the experiments question may be possible in school *in principle*, lab work is actually

[7]"Gymnasium," see Ref. [113] for background about the German school system.

[8]Note while the missing rate (19% at follow-up) is above the value considered as small [117], it is lower than that of Itzek-Greulich *et al.* ([118]; 32% at follow up) or of other studies in this journal (e.g., Henderson *et al.* [113]; 24%). Moreover, participation in all comparison groups took place within regular, compulsory teaching hours; missing was thus rather due to random causes like illness, etc., and not likely to be due to a selective "leaving out" caused by reasons like disinterest, etc.

[9]www.iphysicslab.de (20.06.22).

limited in many schools, due to lack of time (of the teachers), high equipment costs, or poor school facilities [119]. This is true in particular for schools in underprivileged areas [120,121]. Consequently, it is a relevant question to investigate the educational outcomes of such a setting. Second, earlier research with well-controlled studies on this question is scarce, and even the most recent and advanced studies [67,118] suffer from several considerable limitations (different scripts and different educators in the comparison groups, content knowledge test aiming at lower-order learning; very small effect sizes for pre-post learning gains; see Sec. II. D. 1). Thus, there is a need of further research on the "SOL = school" setting which goes beyond the above limitations, as proposed here (see below for control measures and test instruments).

The pretest took place in the physics lesson preceding the intervention. The midintervention test took place immediately after the lab work unit to assess its immediate impact. The post-test was carried out in the physics lesson after the end of the intervention, and the follow-up test after a six-week delay to the post-test. In the time between post-test and follow-up test, standard physics lessons (on another topic, not pressure and buoyancy) took place for all learning groups.

By agreement with the physics teachers involved, it was ensured that all participating classes (i) were approximately in the middle of the teaching unit on buoyancy at the beginning of the intervention and (ii) that they did not receive any physics lessons between pre- and post-test beyond those according to the study design.

To avoid confounding effects, several further variables were kept constant or controlled (see Sec. II. C for background): First, all groups had for all lessons the same instructor who was also the routine instructor at the SOL and at school (an experienced school teacher and the first author of this study); moreover, the perception of instructor engagement by participants was measured to check for comparability across groups. Second, several variables that could not be kept constant over all participants were considered in the analyses: gender, prior knowledge, initial interest and self-concept, language skills, and general intelligence (see Sec. III. D for measures).

C. Instructional environment and material

The iPhysicsLab[9] is a classic science outreach lab that is visited by entire classes as part of half-day events. Its laboratory spaces are also used for normal university teaching. The various modules are usually directly related to the curriculum, so that they can be fully integrated as an experimental supplement in a lesson sequence. Pupils are given the opportunity to conduct and analyze experiments and test hypotheses in autonomous group work, supported by the educator. In the module on fluid pressure and buoyancy investigated in this study, hands-on experiments are used to ensure feasibility at the learning place school (which would not be possible with expensive high-end experiments). The lab work unit of all groups contained four experiments:

- Hydrostatic pressure I: Bursting of a wine barrel according to Blaise Pascal [122].
- Hydrostatic pressure II: Maximum height for drinking through a straw [123].
- Buoyancy I: Stone thrown from a boat—what about the water level? [124,125].
- Buoyancy II: Cartesian diver [126,127].

Participants worked in groups of two, following a circuit through the above four experimental stations.

As an example, Fig. 4 shows the implementation of the experiment Hydrostatic Pressure I as hands-on version. Instead of the historical wine barrel, a commercial Tetra-Pak was used, which can be burst by the hydrostatic pressure of an approximately 4 m high head of water.

The preparation lesson (TG1, CG) addressed the following aspects:

- Providing necessary prior knowledge
- Development of motivating research hypotheses
- Creation of a contextual link between physics lessons at school and the lab work unit at the SOL

In particular, the question "Why do ships float?" was used to address different topics of buoyancy and hydrostatic pressure, providing necessary prior knowledge to carry out the SOL experiments. The specific contexts used in the experiments were not addressed during preparation, except for the experiment "Stone thrown from a boat," which served as a link between the preparation and lab session (Fig. 5). A newspaper article was used to contextualize various questions on buoyancy. A mathematical treatment of the topic according to the curriculum [128] took place in the form of a written homework assignment on the day of the lab work unit. For the follow-up work, a learning quiz about selected contexts was designed. Homework related to the SOL experiments served as another follow-up activity.

For the group without preparation and follow-up activity (TG2), the lessons before and after the SOL visit were standard physics lessons for this stage of the teaching sequence on buoyancy, without specifically relating to the SOL experiments.

D. Measures

1. Interest and self-concept

These tests were based on two well-validated instruments [103,129,130], which were further developed and validated by Kuhn and Müller [131,132]. They consisted of two subscales for (situational) interest (int) and for self-concept (sc) with a total of 18 items and a six-level Likert scale. All test properties were within the recommended range [133,134], see Table III for pretest values (values at other measurement times are very similar). Figure 6 shows various sample items of the scales surveyed.

FIG. 4. Hands-on experiment on the bursting of a wine barrel according to Blaise Pascal [122]. Because of the necessary height difference (several meters), part of the activity was carried out outside the building.

2. *Learning*

To measure students' learning of the subject matter of "pressure and buoyancy" a research-based, curricularly valid test with a focus on conceptual and procedural understanding (cpu) was used. This consisted of a total of 10 items, which included both multiple-choice and free-text items related to procedural understanding of the experiments and known conceptual difficulties for the subject matter ("sucking vacuum," Achimedes' principle, etc. [135–138]; see Table VII in the Appendix for a more detailed specification). Curricular validity and comprehensibility of the test items were ensured by an expert validation.[10] All test properties (Table IV) were within the recommended range for a classroom assessment [133,134]. In that respect, a comment on the value of α_C is in order: Current practice and recommendations

[10]Four experienced teachers of the target age group (mean teaching experience 16.5 yr) were asked on a 6 point Likert scale (1: complete disagreement; 6: complete agreement) about (i) curricular validity; (ii) appropriateness as a physics test question; and (iii) whether they would use a given item in a test. For all questions, agreement of the teachers was very high, with little variation [averages (standard deviations) for all items and teachers: 5.6 (0.5)]; 5.6 (0.5); 5.5 (0.5), for the three questions, respectively.

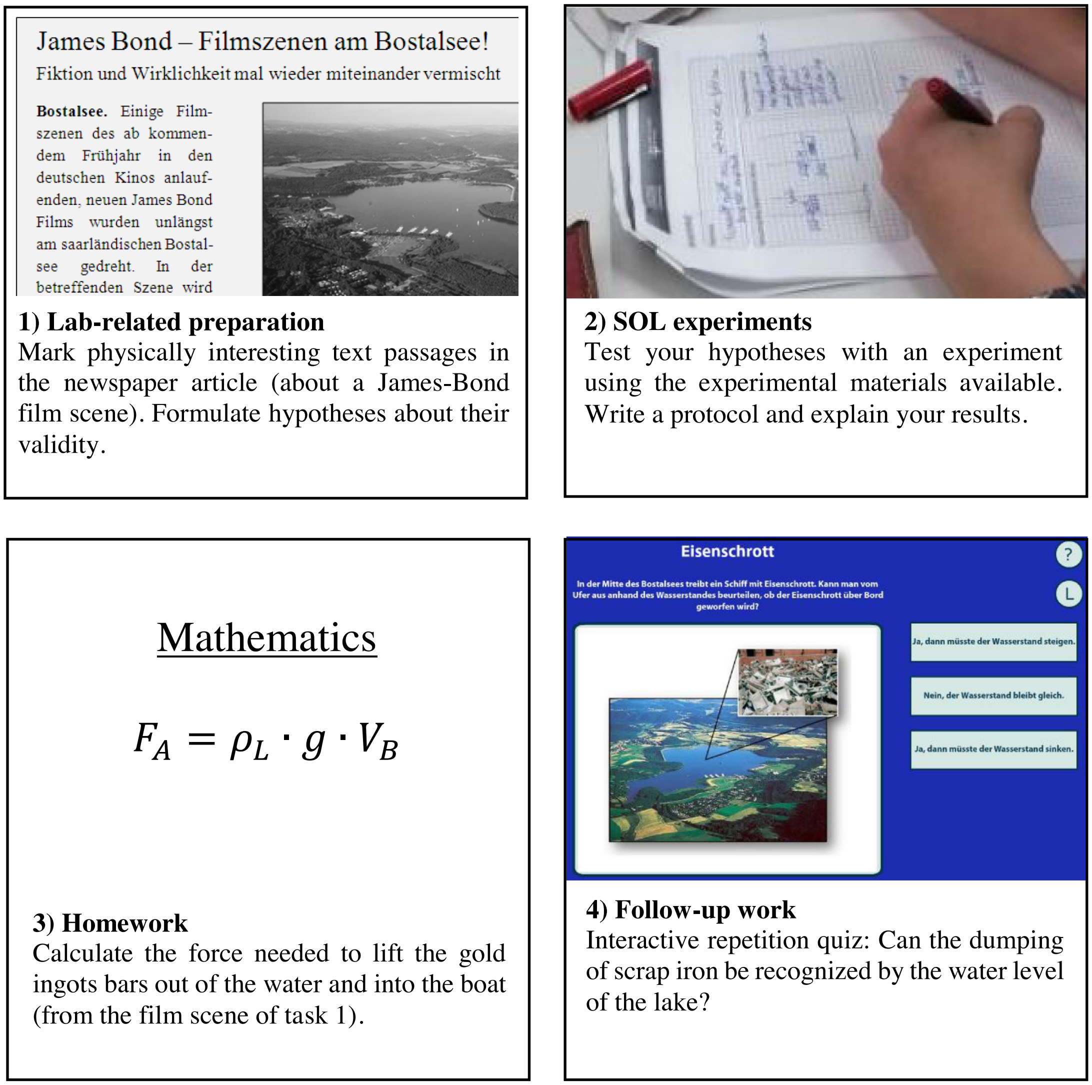

FIG. 5. Instructional material for preparation and follow-up work related to SOL experiment "Stone thrown from a boat".

emphasize that assessment instruments, when covering multiple concepts, as, e.g., in particular in classroom contexts, are expected to have low values of alpha, and that for a test with meaningful content coverage, low internal consistency is not a major impediment to its use [139,140]. Adams and Wieman [141], in their review on assessment instruments in science education, state that instruments designed to measure multiple concepts in a short time (i.e., in a typical classroom setting) are expected to have low values of internal consistency). Indeed, according van Blerkom [142], typical classroom tests will display values between 0.60 and 0.80. Accordingly, for group level measurements, a value of 0.6 falls in the acceptable range [133,143].

TABLE III. Characteristics of the interest or self-concept test (pretest). Mean (M), standard deviation (SD), discrimination index (D), item-test-correlation (r), Ferguson's delta (δ), Cronbach's alpha (α_C) [134]; target range according to Ding and Beichner [134] and EPFA [133].

Property	M	SD	D	r	δ	α_C
Sc	2.8	0.8	0.41	0.49–0.70	0.97	0.90
Int	3.9	0.9	0.48		0.98	0.87
Target range			≥ 0.3	≥ 0.2	≥ 0.9	≥ 0.6

3. Control variables

Several learner characteristics were taken into account (see Sec. II. C). Reading and language competence were

TABLE IV. Characteristics of the learning test (post-test values). Solution probability (P), discrimination index (D), item-test-correlation (r), Ferguson's delta (δ), Cronbach's alpha (α_C) [134].

Property	P	D	r	δ	α_C
Value	0.58	0.46	0.20–0.37	0.93	0.60

Interest	At home, I look in books, on the internet or similar to find out more about topics from physics lessons.	
Self concept	My performance in physics is good according to my own assessment.	
Learning	Michael is not allowed to drink cola at home. His trick to get around the ban: He makes himself a long XXL straw and wants to use it to drink unnoticed from a can of Coke that is 11 metres below him on the floor under his bedroom window. Which statement is correct?	
	☐	His trick works because physically only the diameter of the straw is relevant for this experiment.
	☐	His trick would theoretically work, if only humans could suck enough on the straw.
	☐	His trick doesn't work because it is physically impossible to drink through a straw from that height.
	☐	His trick works because the higher air pressure at the bottom makes the cola in the straw rise to the top.

FIG. 6. Example items from used scales.

assessed using a standardized test instrument ([144],) and by the last German language grades. General (nonverbal) cognitive abilities were also assessed by a standardized test instrument ([145], $\alpha_C = 0.80$). As a measure of initial knowledge, the last physics and mathematics grades and the knowledge test mentioned above were used. Initial interest and self-concept were also assessed (with the instrument mentioned above). Instructor engagement and support across groups, as perceived by participants, was measured by a 5-item scale (instructor perception, IP), where two items were based on those used in other SOLs [41,43], the other items were self-developed. One dimensionality was established by exploratory factor analysis; the internal consistency was $\alpha_C = 0.8$.

E. Analysis

An analysis of covariance (ANCOVA) was carried out to investigate for the main effects of the intervention, and the influence of the control variables (calculation of adjusted values). To determine the source of differences, *post hoc* tests were used. In detail Tukey analysis was employed. The statistical requirements for ANCOVA were examined in advance of the analyses [146,147]. All calculations were performed using the analysis software SPSS 21.[11]

[11]https://www.ibm.com/support/pages/spss-statistics-210-available-download (01.06.22).

Effect sizes are reported as Cohen's d, that is, the group mean difference divided by the pooled standard deviation [56], being very widespread in empirical research and offering the advantage of comparability with other studies. The calculation of effect sizes was based on Wolf [148] and Lipsey and Wilson [55], using adjusted mean values for the group mean difference and the pooled standard deviation.

The Hake index (g) as a measure of the learning gain is reported, which was computed as the ratio of attained to the maximum possible learning gain [149,150].

IV. RESULTS

A. Learning

Figure 7 shows the descriptive data of the learning test at the three test times: pre, post, and follow-up (mean values together with standard errors). The statistical analysis revealed the following results: First, significant learning across groups over the course of the intervention is clearly recognizable [$F(2,292) = 261.56$, $p < 0.001$]. *Post hoc* analyses showed significant learning gains both at the post-intervention and at the follow-up test. Effect sizes for pre-post learning gains were very large for all groups ($d^* = 2.1$, 2.3, 1.0 for CG, TG1, and TG2, respectively). The standardized learning (Hake index g) at the sample level between the pre- and post-test was $g = 0.4$. For the follow-up measurements, gain effect sizes were reduced but still very large for TG1 and CG, and medium-sized for TG2 ($d^* = 1.8$, 1.8, 0.7 for CG, TG1, and TG2, respectively).

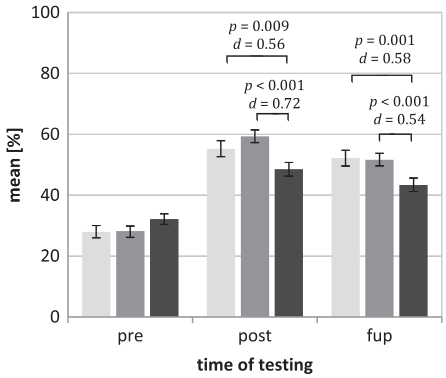

FIG. 7. Descriptive data of the learning test with standard error of mean. Significances and effect sizes refer to the pairwise comparisons of adjusted group means. Abbreviations: prep, preparation; fupa, follow-up activity.

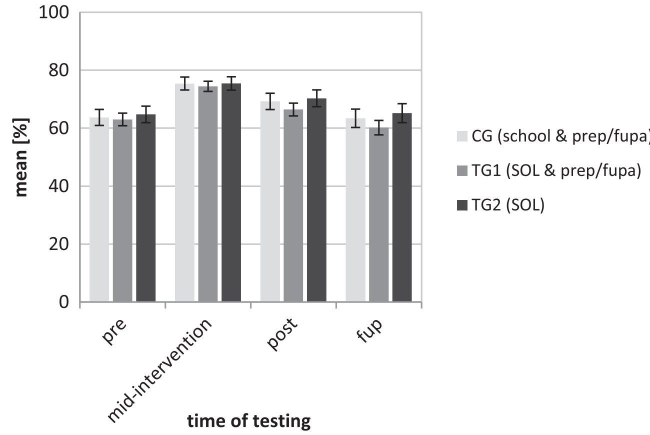

FIG. 8. Descriptive data of the subscale self-concept with standard error of the mean.

Second, the results of the ANCOVA with pairwise *post hoc* tests (Tukey test) revealed significant group differences for learning (see Table V). At both the post and follow-up test, TG1 and CG showed significantly higher scores (cpu_post, cpu_fup) than TG2 [Post: $F(2,153) = 11.94$, $p < 0.001$; Follow-up: $F(2,144) = 7.78$, $p = 0.001$]. Effects sizes were medium or large at the post-test (TG1/TG2: $d = 0.72$; CG/TG2: $d = 0.56$), and medium sized at the follow-up test (TG1/TG2: $d = 0.54$; CG/TG2: $d = 0.58$). TG1 and CG, by contrast, did not differ significantly from each other at either time. Of the covariates surveyed, prior knowledge (cpu_pre), pre-intervention self-concept (sc_pre), prior grade in physics (gr_ph), and reading competence (rc) had an impact on the outcome. They contributed significantly to the explained variance for learning at the post-test (cpu_post).

B. Interest and self-concept

Figures 8 and 9 show the descriptive data of the two subscales of self-concept and interest at group level. The group means of the respective subscale at the four measurement times pre- intervention, midintervention, postintervention, and follow-up are displayed, as well as the standard errors of the mean.

ANCOVA results for both subscales are presented in Table VI. For the midintervention state test directly after the lab sessions, significant increases for both subscales were found [sc: $F(3,351) = 41.83$, $p < 0.001$; int: $F(3,351) = 96.42$, $p < 0.001$]. Effect sizes were medium size ($d^* \approx 0.5$) for self-concept and large (>0.8) for interest. For the post-test measurement time (approx. 1 week after the SOL visit), a significant difference to the prescores was only found for self-concept. The follow-up results of both scales were back to the pre-intervention level. There were no significant between-group differences at any time, except for interest at the follow-up measurement [$F(2,129) = 4.78$, $p = 0.010$, $d = 0.38$]. Of the covariates surveyed, the pre-intervention value of interest had the strongest impact on the same variable at all later measurement times. The same held analogously for self-concept, which also had an impact on

TABLE V. ANCOVA results of the learning test at post-test and follow-up. Dependent variable (DV); Error degrees of freedom (df_e); F statistic (F); observed significance level (p); effect size Cohen $d(d)$; adjusted means (M_a) (abbreviations: cpu_pre: prior conceptual and procedural understanding; sc_pre: pre-intervention self-concept, gr_ph: prior grade in physics; rc: reading competence; fup: follow-up).

				p				d			M_a		
DV	df_e	Factor (CV)	F	Tot	CG/TG1	TG1/TG2	TG2/CG	CG/TG1	TG1/TG2	CG/TG2	CG	TG1	TG2
cpu_post	146	group	11.94	<0.001	n. s.	<0.001	0.009	⋯	0.82	0.52	56.3	59.8	47.9
		cpu_pre	19.55	<0.001									
		gr_ph	10.52	0.001									
		sc_pre	8.45	0.004									
		rc	3.03	0.035									
cpu_fup	137	group	7.78	0.001	n. s.	<0.001	0.001	⋯	0.59	0.59	51.9	52.5	43.5
		cpu_pre	37.71	<0.001									
		gr_ph	9.61	0.002									
		rc	5.19	0.024									

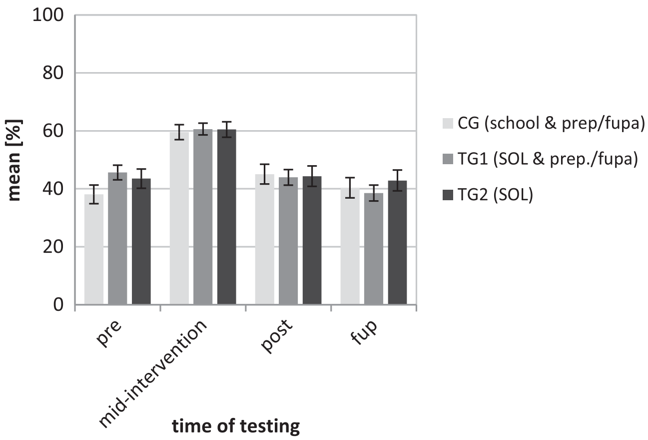

FIG. 9. Descriptive data of the subscale interest with standard error of the mean.

interest as the two measurement times after the lab session but not at the follow-up measurement. Finally, prior grades in mathematics were a predictor of self-concept at the mid-intervention measurement.

C. Gender effects and further results

Both self-concept and interest correlated significantly with the learning at sample level at all three measurement times of the learning test (pre: sc: $r = 0.36^{***}$, int: $r = 0.22^{**}$; post: sc: $r = 0.34^{***}$, int: $r = 0.28^{**}$; follow up: sc: $r = 0.47^{***}$, int: $r = 0.30^{***}$).

Gender analyses at the sample level showed small significant effects at several test times in favor of a higher self-concept of boys [midintervention: $F(1, 137) = 6.24$, $p = 0.014$, $d = 0.38$; post: $F(1, 127) = 9.16$, $p = 0.003$, $d = 0.41$; follow-up: $F(1, 126) = 5.70$, $p = 0.018$, $d = 0.34$]. No such differences were found for interest.

No interactions with the group were detected with regard to any of the control variables collected or at any time of testing, nor any differences in instructor engagement and support across groups.

V. DISCUSSION

A. Research hypotheses on learning (H1, H3a)

The SOL studied in this work has pronounced positive effects on conceptual and procedural understanding, which is reflected in the large group-wide pre-post effect sizes[3] ($d^* = 2.3$ and 1.0 for TG1 and TG2 respectively). These values are in accord with previous research on learning gains in SOLs (pre-post comparison, $d^* \geq 0.8$, see Sec. II. A. 2). Consistently, the Hake gain index found ($g = 0.4$ on the sample level) was in the range found for teaching approaches considered as effective, that is, interactive engagement courses ($g = 0.48 \pm 0.12$; [149,151]). A comparison of the two treatment groups TG1 and TG2 shows that the integration of a science outreach lab in the form of preparation and follow-up activity has a positive effect on students' learning ($d = 0.72$). This effect was still present six weeks after the intervention ($d = 0.54$). Thus, hypotheses Hyp1a and Hyp3a were confirmed. These results are in good agreement with previous studies (see Sec. II. B. 4), although not with Itzek-Greulich *et al.* [67], who found no substantial learning gains, and no advantage of the integrated setting.

TABLE VI. ANCOVA of the self-concept and the interest variables at midintervention, postintervention, and follow-up. Dependent variable (DV); Error degrees of freedom (df_e); F statistic (F); observed significance level (p); effect size Cohen $d(d)$; adjusted means (M_a) (abbreviations: sc_mid: midintervention self-concept; int_pre: pre-intervention interest, gr_m: prior grade in mathematics).

					d	M_a		
DV	df_e	*Factor (CV)*	F	p	TG1/TG2	CG	TG1	TG2
sc_mid	140	group	0.14	n. s.		74.9	75.3	74.2
		(sc_pre)	75.12	<0.001				
		(gr_m)	5.94	0.016				
sc_post	129	group	0.38	n. s.		68.7	67.0	68.8
		(sc_pre)	155.37	<0.001				
sc_fup	130	group	1.54	n. s.		64.3	61.0	64.8
		(sc_pre)	172.78	<0.001				
int_mid	140	group	0.06	n. s.		61.5	60.8	60.9
		(int_pre)	19.96	<0.001				
		(sc_pre)	3.77	0.050				
int_post	130	group	1.55	n. s.		47.8	42.4	44.0
		(int_pre)	19.95	<0.001				
		(sc_pre)	11.35	0.001				
int_fup	129	group	4.78	0.010	0.38	43.6	36.6	43.8
		(int_pre)	54.41	<0.001				

The comparison between TG1 and CG shows that the learning place SOL *per se* does not lead to increased learning. Therefore research hypothesis Hyp1b was rejected. This result is in agreement with that of Itzek-Greulich *et al.* [67].

Itzek-Greulich, too, was unable to prove a connection between the learning place and learning gains. Other work had found positive effects of out-of-school learning places as such (see Sec. II. A. 2). Our interpretation would be that for well-prepared lab work units (with identical content and implementation), where strong learning gains occur also at school, and with a well-controlled comparison, in particular also taking account of initial knowledge, little evidence exists for the "place hypothesis," that is, superior learning taking place at an SOL.

Taken together, these findings show that SOLs can indeed be designed to be effective for learning: An integrated setting with preparation and follow-up activity can have large learning gains and be as effective as learning at school. However, this is not the case for SOLs without integration.

B. Research hypotheses on interest and self-concept (H2, H3b)

A group-wide, short-term increase of interest and self-concept was observed directly after the lab sessions, with medium-size and large effect sizes, respectively (IE: $d^* > 0.8$, SC: $d^* \approx 0.5$). It had already disappeared for the most part one week later (post-test). At the follow-up time, both scales were back to their pre-intervention levels. This confirms short-term catch effects as described by previous studies (Sec. II. B. 2). The interest and self-concept level measured at midintervention time is also in line with other studies, which reflects well on the comparability of the SOL evaluated here with others. There were no group differences at any time of testing, except for self-concept at the follow-up measurement (TG2 > TG1). As there were no effects subsequent to the intervention, and since new teaching topics were started in the time between post- and follow-up test by all participating learning groups, it is unlikely that the effect in question is related to the intervention.

Thus, students' interest and self-concept in this SOL turned out to be independent both of the integration of the SOL (TG1) and of the out-of-school location at a university (TG2). Research hypotheses Hyp2 (a and b) were therefore rejected: no evidence could be found to support the hypothesis often formulated in literature that the connection between short-term "catch" and medium- to long-term "hold" effects for interest is fostered by the integration of SOLs in pre- and post-learning opportunities at school [39–41]. Rather, our findings are consistent with other studies reporting that integration does not have the desired effect on affective variables [43,118,152].

On the one hand, these results raise the question of criteria for effective forms of integration of SOLs, on the other hand, the question of possible dose-effect relationships. Such questions have the potential to increase the effectiveness of learning of SOLs and should therefore be the focus of future research.

C. Gender, other learner characteristics, and further results

No differences were found between girls and boys for the gains in learning and situational interest in any of the settings: the experimental units appeared suitable for both genders both at school and at the SOL. Priemer *et al.* [81] in a large-scale study ($N > 10000$) on situational interest after SOL visits also found no differences between girls and boys.

Self-concept was also increased for both genders ($d^* \approx 0.5$), although more for boys than for girls ($d \approx 0.4$). Thus, our study did not show the small decrease of the "gender gap" (from $d = 0.9$ to 0.7) for self-concept reported by Euler [49], but with smaller differences between boys and girls altogether (d in Euler [49] twice as large than ours). No further inferences about the gender gap can be drawn on the basis of the present data.

From the other learner characteristics, prior knowledge and prior grade in physics had a significant impact on learning, consistent with existing evidence (Sec. II. C). Moreover, reading competence also contributes to the explained variance of the learning variable. This shows the importance of text comprehension for science learning, especially if instruction materials contain larger text passages [153]. Lastly, pre-intervention self-concept also had an influence on the learning outcome, consistent with previous research (Sec. II. C).

As for learning, the pre-intervention scores of the collected affective variables were also the strongest predictors of the respective scores after the SOL visit. This is plausible, as both interest and self-concept, are stable variables. Further, prior grades in mathematics were predictors of self-concept at the midintervention measurement. Again, it is plausible that grades influence academic self-concept, but this is an individual result, not generalizing from mathematics to physics, nor to all measurement times. As the main reason for including this variable was to control for potential differences between the intervention group, a more systematic investigation of these influences is beyond the scope of the present study.

Note that both for learning and affective outcomes and for any of the above predictors, including gender, there were no interactions with the intervention: lab work at school, SOLs, and their enhancement by an integrated setting work equally well for learners of all kinds, in particular for boys and girls and for low and high achievers.

Finally, the correlations found for learning with interest and self-concept are in line with previous research. For interest, the correlation with learning was in the range of $0.22 \leq r \leq 0.30$ at all times of testing, consistent with meta-analytic results ([106]: $r = 0.3$). For the specific case of SOLs, Glowinski [41] reported somewhat higher values

($0.3 \leq r \leq 0.6$) for various components of interest and knowledge; however, that she considered the correlation to subjective self-assessment of learning, while in the present work, a measure by an actual knowledge test was used. For self-concept, the correlation with learning was in the range of $0.34 \leq r \leq 0.47$ at all times of testing, in line with previous research showing that confidence in one's own competence in a specific field or for a given task is a strong predictor of academic effort and achievement [154]. Hattie [57] states that "the relation between self-efficacy and achievement […] is among the strongest of self-measures" (meta-analytic value $d = 0.43$).

D. Limitations and future research

The present study is limited to a given age group (secondary level I), and academic-track schools within the German school system.[7] Our findings about SOLs and the positive effects of integrated approaches (pre- and post-visit activities at school) would have to be confirmed for other educational settings, for example, for other topics (e.g., modern science) and learner groups (e.g., with lower academic abilities). However, no interaction was found for any of the outcomes (cognitive and affective) and for any of the control variables (previous knowledge in physics and mathematics, language, and reading competences, etc.; see previous section). Thus, even though these predictors (e.g., previous knowledge) might be different in other settings, they should only have a moderate influence on the educational advantages of learning at SOLs.

The following limitations on the methodological level have to be mentioned. First, in the present design, only two main features were studied (SOL vs school and integrated setting vs non-integrated setting). Thus, further investigations should study additional factors and their interactions in a more detailed way, such as different forms of integration providing other forms of preparation and orientation. An interesting and practically relevant question is whether background information about an SOL visit (schedule, location, other circumstances [93]) has an orienting, positive effect on educational outcomes beyond the cognitive preparation studied here. A promising conceptual framework for such more fine-graded investigations is provided by the "novelty space" theory of Orion *et al.* [63,95,155], which differentiates different dimensions of novelty, such as cognitive novelty, novelty related to the setting of an out-of-school learning offer, etc. This could add to a more complete evidence-based understanding of "best practice" and success factors for the design and operation of science outreach offers, of obvious interest to their practitioners and operating institutions [156].

Second, the present study has considered only a limited set of outcome variables, which should be broadened by future work. A relevant example in that context is curiosity, which is an important variable at the affective-cognitive intersection [157], and belongs to their very rationale of science outreach offers, as seen by research, scientists, and providers alike [3,158–160]. Third, the sample size in this study is smaller than, for example, that of the study by Itzek-Greulich *et al.* [67], and small main or interaction effects of interest might need larger sample sizes to attain statistical significance. However, an important control measure was realized in this study, as the educator was the same for all treatment groups and the entire sample, which imposed a limit to a manageable sample size of a few hundred.

Finally, while this study found learning effects beyond the short term, it did not provide evidence for motivational "hold effects" expected on the basis of previous research (Sec. II. B. 2). A possible hypothesis is that this needs a post-visit integration lasting longer than a single follow-up lesson. The design of the present study did not allow us to answer this question, which appears to be an interesting topic for future research on SOLs.

VI. CONCLUSIONS AND PERSPECTIVES

To optimize the attitudinal and cognitive gains, teachers and their helpers have to provide support and "scaffolding" between the pupils' existing concepts and the exhibits. Consequently, the quality of classroom preparation, activities within a science center, and follow-up activities are all important aspects of visits [161].

A. General conclusions

The objective of this intervention study was to investigate in detail the effectiveness of SOLs in terms of their integration with teaching at school in the form of preparation and post-visit activities and their effectiveness as an out-of-school learning place. The investigation was carried out on the basis of a SOL thematic unit on "pressure and buoyancy," and it was aligned with a series of methodological exhortations raised for the field of OSLOs (see Sec. II. D. 1) concerning design, instruments, consideration of control variables, underreporting, and including more than short term effects. Its findings confirm and add to the understanding of out-of-school experiences for science learning in the following way.

The study goes beyond recent research [32,67] by investigating a setting where sizable learning actually occurred, using a research-based, curriculum-valid measure of learning, and extended control measures (same lesson script and same instructor plus measure of instructor engagement and support). Finally, in providing evidence for learning effects beyond the short term, the present study contributes to filling the research gap in this area (see Secs. I and II. D. 1) and to put into practice recurrent recommendations for improving research and practice in out-of-school science offers [3,159].

The findings about SOLs without integration are as follows: Even though significant learning and affective pre-post gains were also found for SOL without integration, the school lab work unit was superior for all outcomes: lab work at school, which according to current best practice and evidence includes preparation and follow-up work, is more effective than an SOL visit with inadequate preparation and follow-up activities. Thus, a mere change from the learning place "school" to the learning place "science outreach lab" ("place hypothesis") does not yield superior effects on students' self-concept and interest nor learning, provided that content and implementation do not differ between learning places. Note that this applies to settings where experiments at the SOL and in school are identical (see Sec. III. B for the rationale of these settings). While there are of course SOLs with activities well beyond what is possible in school, there are still good reasons for considering SOLs with experiments identical to school activities, to which we come back in Sec. VI. C.

Even though significant learning and affective pre-post gains were also found for SOL without integration, the school lab work unit was superior for all outcomes: A lab work unit at school, which according to current best practice and evidence includes preparation and follow-up work, is more effective than an SOL visit with inadequate preparation and follow-up activities. Thus, a mere change from the learning place "school" to the learning place "science outreach lab" ("place hypothesis") does not foster students' self-concept and interest nor their learning, provided that content and implementation do not differ between learning places. However, there are still good reasons for considering SOLs with experiments identical to school activities, to which we come back in Sec. VI. C.

Furthermore, the integration of SOLs into physics teaching at school by preparatory and postprocessing teaching units is promising. No negative effects on affective variables were found, but a moderate to large effect on learning ($d = 0.8$), which was still present six weeks after the intervention ($d = 0.6$). Taken together, the available results showed that an SOL with an integrated setting was conducive to learning, without negative side effects on its motivational outcomes. A possible drawback that integration of SOLs with work at school gives a too "school-like" flavor to the visit, and that, in particular, preparation might eliminate its "suspense" and thus lead to a loss of its motivational potential, does not appear justified. With regard to short-term motivational "catch" effects found in similar studies, the present intervention study confirmed the current state of understanding of out-of-school learning offers. It does not, however, provide evidence of medium- to long-term motivational "hold" effects due to the integration of SOLs; as stated above, this question remains inconclusive at the present stage of research, and may be better understood with a corresponding design in future studies.

The insights gained into the cognitive effects of integration are particularly important in view of the fact that around two-thirds of teachers currently neither prepare nor post-process an SOL visit in their lessons [43,162]. The majority of them, in connection with attending an out-of-school learning place, only pay attention to existing links of the topics offered to the school curriculum, the possibility of performing experiments, and the costs of the visit [71]. Here, the results of the study should encourage teachers and SOLs to rethink their approach in order to generate lasting effects on learning for the participating students.

In sum, we see the added value of the present contribution as two-fold: adding evidence for the SOL = school format actually existing, in particular regarding the beneficial effects of integrated settings, and going beyond substantial limitations in prior research; and adding useful perspectives for classroom practice and future development of SOL in more general settings to which we turn now.

B. Practical implications

For outreach practitioners and operating institutions, the following findings appear of interest. First, the present study provides new evidence, responding to the requirement of improved methodology in the field (see Sec. II. D. 1), that substantial learning at SOLs is possible, with positive effects lasting beyond the short term. There is thus no need to exclude cognitive outcomes from the objectives of science outreach offers. Moreover, the present findings add to existing evidence about the substantial motivational potential of SOLs. Second, integration with classroom teaching was shown to be necessary for learning, and not harmful for motivation. A disconnected SOL experience is not likely to spark learning (and all the less in a lasting way), well in accord with relevant research about learning processes.

Third, the effect sizes of learning and motivational effects as measures of practical importance corroborate the educational potential of SOLs: very large ($d^* > 0.8$) for learning and interest, and medium sized for self-concept ($d^* \approx 0.5$)[3]. Compared to Hattie [57], the effect size for learning of an SOL with vs without integration ($d = 0.72$ at post-test) would come out on position 13 of 138 instructional approaches. Although this is an individual study, not a meta-analysis, the size for integrated SOLs appears to be in the interesting range, and it was obtained by only one lesson of preparation and of follow-up activities (which would be included in classroom teaching anyway).

Fourth, the fact that there are no interactions of different covariates with the outcome variables is also of practical interest. This means that the enhancement of SOLs by integration works for diverse kinds of learners, and that in particular it is not restricted to learners of a higher initial level of motivation or understanding.

Fifth, when compared to lab work at school, the effects of an SOL (with integration) on learning and motivation are as good as those in a well-prepared classroom setting, but not better. In view of the substantial absolute gains (see previous point), this is certainly not bad; however, it leads to an obvious question: What then is the added value of SOLs?

C. The added value of science outreach labs, and perspectives for future research

In the following, we present several arguments in favor of SOLs as valuable components of physics education. To begin with, we discuss the question of whether SOLs with experiments identical to school activities have an added value. First, while SOL experiments may be possible in school *in principle*, lab work is actually often limited in schools, both by constraints of budget and time. In particular, with the ever-increasing charges of everyday work in school, this often leaves little time for teachers to realize experimental offers for learners, and even less to develop new lab work at a level possible in an SOL, even if this would be possible in principle. In this case, high-quality lab work provided by SOLs is a substantial support for teachers. Moreover, if committed and creative scientists at a university realize the development of stimulating activities which *also* could be carried out at school in terms of necessary equipment (e.g., low-cost experiments), this could improve physics education even further, as teachers will learn these ideas and transfer them to their classroom practice (in fact, this is a major reason for them to visit SOLs [119]). Another way of such a "spin-off" of SOLs can be the development of experimental kits, which even can help to solve the budget or equipment problem at some schools. In both cases, our results show that educational outcomes at schools then can be comparable, which we find encouraging and relevant. Finally, such offers are even more valuable for schools in underprivileged settings: the effects on cognitive and affective variables by lab work at school can be equal to those by a well-equipped SOL, provided that the necessary material, staff, and time resources are available. If this is not the case, SOLs can ensure a sizable contribution to educational equity.

Second, SOLs and other science outreach offers can of course provide experimental activities not possible in schools, for example, for topics of modern science (where the equipment is not available, see, for example, the SOLs at CERN [11], the Paul-Scherrer-Institute [34], or Fermilab [13], for the dimensions, degree of sophistication or other special features of the experiments [10,14,163,164] or for practical reasons (maintenance, security, etc.). In this context, we fully agree with Braund and Reiss' [7] statement about "*extended and authentic practical work"* and *"access to rare material"* as important elements of science outreach offers. Here, the outcomes of our study add to the discussion about SOLs two perspectives we find significant: In the broad range of SOLs formats, there are offers which downplay learning objectives, with the argument that these would create a "didactic" atmosphere doing harm to the motivational objectives. Our study shows that both objectives can well coexist. Similarly, there are offers which follow rather an "edutainment" approach which often avoids an integration with pre- and post-activities at school, with the same argument of a didactic atmosphere (see also Sec. VI. A). Our study shows the contrary: a well-thought integration can *increase* motivational effects (here interest and self-concept). Of course, this remains to be shown for offers different from a "SOL $\neq$ school" setting studied in the present work (Sec. III. B), and we see this as an interesting and relevant question for future research.

Third, there are important objectives of SOLs that go beyond motivation and learning a specific topic. A visit to an out-of-school learning offer at a research institute can provide experience and insight regarding the work done in a genuine research environment, and its purposes and societal value, not (or less) possible by other offers [1,3,7,8]. In particular, personal contact with researchers provides opportunities to experience authentic ways of doing science (as opposed to the restricted or even artificial views on science as they may sometimes be presented in school), and to overcome misconceptions and stereotypes related to scientists as person and science as professional activity. Moreover, the contact with scientists may convey the experience of curiosity, excitement, and satisfaction found in her work [8,83]. Tytler *et al.* [165] argue that such encounters can be a valuable contribution to providing role models of adults interested in and committed to science and related areas. Again, it is an interesting question for further research how these additional features combine with integration (which was the factor studied in the present work).

In conclusion, our answer to the question "From enjoyment to learning?" (Sec. I) is "Yes, this is possible!" and integration of SOLs is a key to this. We consider SOLs as a promising element of science education, both with experimental activities possible also in school, and with features going beyond this, well worth considering on the practical level, and for further research for a more complete understanding of their outcomes and success factors.

ACKNOWLEDGMENTS

Financial support by the Ingrid-und-Wilfried-Kuhn Foundation for Physics Education (Munich, Germany) is gratefully acknowledged, as is work for preparing the manuscript by Alexander Brown (St. Genis, France).

A. Molz carried out the study and the data analysis, and wrote the manuscript together with A. Müller and input by J. K.; A. Müller was involved in the design, analysis, and discussion of the study, and wrote the manuscript together with A. Molz; J. K. was involved in the design, analysis, and discussion of the study, and provided input and feedback to the manuscript.

APPENDIX: KEY CONCEPTS OF THE LEARNING TEST

The learning test is about basic conceptual understanding, in particular about several conceptual difficulties of the subject matter "pressure and buoyancy" described in the literature, as specified in Table VII.

TABLE VII. Main concepts and conceptual difficulties of the learning test.

Conceptual difficulty	References
Drinking straw: "sucking vacuum" vs. atmospheric pressure difference	[136]
Forces on a stone (or similar object) in water	[125,166]
Archimedes principle and difficulties related to it	[135,167–169]
Connection of buoyancy and hydrostatic pressure	[169]
Melting ice cube, water displacement	[170]; Chap. "Archimedes Principle
Explanation of the Cartesian diver	[138,168]

[1] L. J. Rennie, Measuring affective outcomes from a visit to a science education centre, Res. Sci. Educ. **24**, 261 (1994).

[2] M. Singh, Why Recognition Matters—Valuing Informal and Non-formal Learning Across the North-South Divide, UNESCO Institute of Lifelong Learning (Springer, Cham, 2009).

[3] National Research Council, *Learning Science in Informal Environments: People. Places, and Pursuits* (National Academies Press, Washington, DC, 2009).

[4] P. Werquin, *Recognising Nonformal and Informal Learning Outcomes, Policies and Practices: Outcomes, Policies and Practices* (OECD Publishing, Paris, 2010), Vol. 2009.

[5] D. Corrigan, C. Buntting, A. Jones, and J. Loughran, *Navigating the Changing Landscape of Formal and Informal Science Learning Opportunities* (Springer, Cham, 2018).

[6] A. J. Friedman, The evolution of the science museum, Phys. Today **63**, 10, 45 (2010).

[7] M. Braund and M. Reiss, Towards a more authentic science curriculum: The contribution of out-of-school learning, Int. J. Sci. Educ. **28**, 1373 (2006).

[8] S. M. Stocklmayer, L. J. Rennie, and J. K. Gilbert, The roles of the formal and informal sectors in the provision of effective science education, Studies Sci. Educ. **46**, 1 (2010).

[9] C. L. Thomas, Assessing high school student learning on science outreach lab activities, J. Chem. Educ. **89**, 1259 (2012).

[10] Heureka (2020), https://www.heureka.fi/kouluohjelmat-2/?lang=en.

[11] CERN, "S'CoolLab," CERN (2020), https://scoollab.web.cern.ch/.

[12] I. Glowinski and H. Bayrhuber, Student labs on a university campus as a type of out-of-school learning environment: Assessing the potential to promote students' interest in science, Int. J. Environ. Sci. Educ. **6**, 371 (2011).

[13] Fermilab (2020), https://ed.fnal.gov/lsc/.

[14] Exploratorium (2020),https://www.exploratorium.edu/.

[15] E. del Barco, PhYSICOS—A Physics Summer Camp for High School students (2017).

[16] APS, *Forum on Outreach, and Engaging the Public of the American Physical Society* (American Physical Society, Ridge, 2020).

[17] D. A. Roberts, Scientific Literacy/Science Literacy, *Handbook of Research on Science Education*, edited by S. K. Abell and N. H. Lederman (Lawrence Erlbaum, Mahwah, NJ, 2007), pp. 729–780.

[18] E. Brewe, V. Sawatelle, L. H. Kramer, G. E. O'Brien, I. Rodrigues, and P. Pamela, Toward equity through participation in Modeling Instruction in introductory university physics, Phys. Rev. ST Phys. Educ. Res. **6**, 010106 (2010).

[19] S. Hyater-Adams, C. Fracciola, N. Finkelstein, and K. Hinko, Critical look at physics identity: An operationalized framework for examining race and physics identity, Phys. Rev. Phys. Educ. Res. **14**, 010132 (2018).

[20] L. J. Rennie, *"A very valuable partnership" Evaluation of the Scientists in Schools Project 2011-12* (CSIRO Education, Canberra, 2012).

[21] C. Rop, *Close to home: A Review of the Literature on Learning in Schoolyards and Nearby Natural Settings, 1980 to the Present* (Urban Affairs Center, University of Toledo, Toledo, OH, 2004).

[22] J. De Witt and M. Storcksdieck, Excursions, *Encyclopedia of Science Education* (Dordrecht, Springer, 2015).

[23] J. H. Falk and J. E. Heimlich, Who is the free-choice environmental education learner?, *Free-Choice Learning and the Environment* (AltaMira Press, Lenham, 2009).

[24] H. Eshach, Bridging in-school and out-of-school learning: Formal, non-formal, and informal education, J. Sci. Educ. Technol. **16**, 171 (2007).

[25] M. Rutter, *Fifteen Thousand Hours: Secondary Schools and their Effects on Children* (Paul Chapman, London, 1994).

[26] S. Hidi and K. Renninger, The four-phase model of interest development, Educ. Psychol. **41**, 111 (2006).

[27] L. JungMi, Cumulative learning, *Encyclopedia of the Sciences of Learning* (Springer, Boston, 2012), pp. 887–893.

[28] L. J. Rennie, Learning science outside of school, *Handbook of Research on Science Education* (Lawrence Erlbaum, Mahwah, NJ, 2007).

[29] A. M. Lucas, Scientific literacy and informal learning, Studies Sci. Educ. **10**, 1 (1983).

[30] K. M. Ellenbogen and R. Stevens, Informal Science Learning Environments: A Review of Research to Inform K-8 Schooling (National Academies of Sciences' Committee on Science Learning K-8, Washington, DC, 2005).

[31] S. Swarat, A. Ortony, and W. Revelle, Activity matters: Understanding student interest in school science, J. Res. Sci. Teach. **49**, 515 (2012).

[32] H. Itzek-Greulich and C. Vollmer, Emotional and motivational outcomes of lab work in the secondary intermediate track: The contribution of a science center outreachlab, J. Res. Sci. Teach. **54**, 3 (2017).

[33] M. Besse, S. Fragnière, A. Mueller, M. Piguet, L. Dubois, D. Miéville, and D. Schumacher, Learning about archaeology and prehistoric life. The effects of two workshops in primary education, Sci. Educ. **28**, 759 (2019).

[34] Paul-Scherrer-Institute, "iLab" (2020). https://www.psi.ch/en.

[35] C. Schroeder, T. Scott, H. Tolson, T. Hunag, and Y. Lee, A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States, J. Res. Sci. Teach. **44**, 1436 (2007).

[36] K. J. Topping, C. Buchs, D. Duran, and K. H. Van, *Effective Peer Learning: From Principles to Practical Implementation* (Routledge, Taylor & Francis Group, New York, 2017).

[37] T. Tal, Out-of-school: Learning experiences, teaching and students' learning, *Second International Handbook of Science Education* (Springer, Dortrecht, 2012).

[38] P. Cohen, J. Cohen, L. S. Aiken, and S. G. West, The problem of units and the circumstance for POMP, Multivariate Behav. Res. **34**, 315 (1999).

[39] K. Engeln, Schülerlabors: Authentische, aktivierende Lernumgebungen als Möglichkeit, *Interesse an Naturwissenschaften und Technik zu wecken* (Logos, Berlin, 2004).

[40] A. Brandt, *Förderung von Motivation und Interesse durch außerschulische Experimentierlabors* (Cuvillier Verlag, Göttingen, 2005).

[41] I. Glowinski, *Schülerlabore im Themenbereich Molekularbiologie als Interesse fördernde Lernumgebung* (Christian-Albrechts-Universität, Kiel, 2007).

[42] CitT, Caught in the trap—an interactive exposition on particle traps and laser optics at the Palais de la Découverte (Paris), in *Demonstrationsexperimente zur Laserkühlung und -speicherung von Atomen* (Albert-Ludwigs-Universität, Freiburg, 2008).

[43] C. Pawek, *Schülerlabore als interessefördernde außerschulische Lernumgebungen für Schülerinnen und Schüler aus der Mittel- und Oberstufe* (Christian-Albrechts-Universität, Kiel, 2009).

[44] W. Zehren, *Forschendes Experimentieren im Schülerlabor* (Universität, Saarbrücken, 2009).

[45] A. Molz, *Verbindung von Schülerlabor und Schulunterricht—Auswirkungen auf Motivation und Kognition im Fach Physik* (Verlag Dr. Hut, München, 2016).

[46] R. Cors, *Informal Science Learning: An investigation of how novelty and motivation affect interest development at a mobile laboratory* (University of Geneva, Geneva, 2016).

[47] F. Simon, *Der Einfluss von Betreuung und Betreuenden auf die Wirksamkeit von Schülerlaborbesuchen* (Logos Verlag, Berlin, 2019).

[48] J. Woithe, *Designing, measuring and modelling the impact of the hands-on particle physics learning laboratory S'Cool LAB at CERN* (CERN/Technical University, Geneva/Kaiserslautern, 2020).

[49] M. Euler, Neugier, Kreativität, Kompetenzförderung. Schülerlabore in Deutschland. Erfolgsgeschichte & Herausforderungen (2007), http://www.dlr.de/schoollab/desktopdefault.aspx/tabid-4507/7380_read-8441.

[50] T. Sikora and A. Pokropek, Gender segregation of adolescent science career plans in 50 countries, Sci. Educ. **96**, 234 (2012).

[51] I. Fuller, S. Edmondson, D. France, D. Higgitt, and I. Rattinen, International perspectives on the effectiveness of geography fieldwork for learning, J. Geog. Higher Educ. **30**, 89 (2006).

[52] M. Rickinson, J. Dillon, K. Teamey, M. Morris, and M. J. Choi, *A Review of Research on Outdoor Learning* (Field Studies Council, Preston Montford, 2004).

[53] J. Maskall and A. Stokes, *Designing Effective Fieldwork for the Environmental and Natural Sciences* (Higher Education Academy Subject Centre for Geography, Earth and Environmental Sciences, Plymouth, 2008).

[54] J. P. Prather, Review of the value of field trips in science instruction, J. Elem. Sci. Educ. **1**, 10 (1989).

[55] M. Lipsey and D. Wilson, *Practical Meta-Analysis*, Applied Social Research Methods Series Vol. 49 (Sage, Thousand Oaks, CA, 2003), Appendix B.

[56] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Lawrence Erlbaum, Hillsdale, New Jersey, 1988).

[57] J. Hattie, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (Routledge, New York, 2009).

[58] J. H. Falk and J. Balling, The field trip millieu: Learning and behavior as a funktion of contextuel events, J. Educ. Res. **76**, 22 (1982).

[59] W. W. Martin, J. H. Falk, and J. D. Balling, Environmental effects on learning: The outdoor field trip, Sci. Educ. **65**, 301 (1981).

[60] J. Falk, W. Martin, and J. Balling, The novel field trip phenomena: Adjustments to novel settings interferes with task learning, J. Res. Sci. Teach. **15**, 127 (1978).

[61] M. Lisowski and J. F. Disinger, The effect of field-based instruction on student understandings of ecological concepts, J. Environ. Educ. **23**, 19 (1991).

[62] W. Killermann, Research into biology teaching methods, J. Biol. Educ. **33**, 4 (1998).

[63] N. Orion and A. Hofstein, Factors that influence learning during a scientific field trip in a natural environment, J. Res. Sci. Teach. **31**, 1097 (1994).

[64] B. D. Brazell and S. Espinoza, Meta-analysis of planetarium efficacy research, Astron. Educ. Rev. **8** (2009).

[65] K. Damerau, Molekulare und Zell-Biologie im Schülerlabor, Doctoral Disertation, Uninversität Wuppertal, 2018.

[66] M. Hirth, *Akustische Untersuchungen mit dem Smartphone und Tablet-Computern: Fachliche und didaktische Aspekte* (Dr. Hut, Kaiserslautern, 2019).

[67] H. Itzek-Greulich, B. Flunger, C. Vollmer, B. Nagengast, M. Rehm, and U. Trautwein, 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 Instr. **38**, 43 (2015).
[68] S. Chapman, L. Catala, J.-C. Mauduit, K. Govender, and J. Louw-Potgieter, Monitoring and evaluating astronomy outreach programmes: Challenges and solutions, South African J. Sci. **111**, 112 (2015).
[69] I. Glowinski and H. Bayrhuber, Student labs on a university campus as a type of out-of-school learning environment: Assessing the potential to promote students' interest in science, Int. J. Environ. Sci. Educ. **6**, 371 (2011).
[70] O. Fallik, S. Rosenfeld, and B. S. Eylon, School and out-of-school science: A model for bridging the gap, Studies Sci. Educ. **49**, 69 (2013).
[71] D. Anderson and Z. Zhang, Teacher perceptions of field-trip planning and implementation, Visitor Studies Today **6**, 6 (2003).
[72] J. Kisiel, Understanding elementary teacher motivations for science fieldtrips, Sci. Educ. **89**, 936 (2005).
[73] I. Schmidt, D. S. Di Fuccia, and B. Ralle, Außerschulische Lernstandorte: Erwartungen, Erfahrungen und Wirkungen aus der Sicht von Lehrkräften und Schulleitungen, Mathematisch-Naturwissenschaftlicher Unterricht **63**, 362 (2011).
[74] M. Storcksdieck, Teachers' perceptions and practice surrounding a field trip to a planetarium, in *Meeting of the National Association for Research in Science Teaching* (NARST, Vancouver, 2004).
[75] R. Watson, The role of practical work, in *Good Practice in Science Teaching: What Research has to Say* (Open University Press, Buckingham, 2000).
[76] M. Tesch and R. Duit, Experimentieren im Physikunterricht—Ergebnisse einer Videostudie, Z. Naturwissenschaften **10**, 51 (2004).
[77] V. N. Lunetta, A. Hofstein, and M. P. Clough, Learning and teaching in the school science laboratory: An analysis of research, theoryand practice, *Handbook of Research on Science Education* (Lawrence Erlbaum, Mawah, NJ, 2007), pp. 393–441.
[78] J. Börlin and P. Labudde, Practical work in physics instruction: An opportunity to learn?, in *Quality of Instruction in Physics. Comparing Finland, Switzerland and Germany* (Münster, Waxmann, 2014), pp. 111–128.
[79] A. Krapp, Interest and human development: An educational-psychological perspective, *Development and Motivation* (The British Psychological Society, 2003), pp. 57–84.
[80] A. Krapp and M. Prenzel, Research on interest in science: Theories, methods, and findings, Int. J. Sci. Educ. **33**, 27 (2011).
[81] B. Priemer, C. Menzl, f. Hagos, W. Musold, and J. Schulz, Das situationale epistemische Interesse an physikalischen Themen von Mädchen und Jungen nach dem Besuch eines Schülerlabors, Z. Naturwissenschaften **24**, 279 (2018).
[82] A. Krapp, Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective, Learning Instr. **12**, 383 (2002).
[83] M. Euler, Hands-on science and informal learning: Challenge and potentials of authentic lab activities, *Informal Learning and Public Understanding of Physics*, edited by G. Planinšič and A. Mohorič (GIREP/Groupe international de recherche sur l'enseignement de la physique, Antwerp (Belgium), 2005), pp. 9–17.
[84] R. Cors, A. Müller, and N. Robin, Towards a more comprehensive theoretical framework for investigating novelty at out-of-school learning places for science and technology learning, Prog. Sci. Educ. **1** (2017).
[85] D. Anderson, K. B. Lucas, I. S. Ginns, and L. D. Dierking, Development of knowledge about electricity and magnetism during a visit to a science museum and related post-visit activities, Sci. Educ. **84**, 658 (2000).
[86] D. Anderson, K. B. Lucas, and I. S. Ginns, Theoretical perspectives on learning in an informal setting, J. Res. Sci. Teach. **40**, 177 (2003).
[87] Smithsonian Institution Office of Policy, and Analysis, *Enhancing the visits of middle-school tour groups to the Smithsonian* (Smithsonian Institution, Washington, DC, 2007).
[88] M. Rollnick, S. Zwane, M. Staskun, S. Lotz, and G. Green, Improving pre-laboratory preparation of first year university chemistry students, Int. J. Sci. Educ. **23**, 1053 (2001).
[89] L. B. Bruck and M. H. Towns, Preparing students to benefit from inquiry-based activities in the chemistry laboratory: guidelines and suggestions, J. Chem. Educ. **86**, 820 (2009).
[90] H. Y. Augustian and M. K. Seery, Reasserting the role of pre-laboratory activities in chemistry education: a proposed framework for their design, Chem. Educ. Res. Pract. **18**, 518 (2017).
[91] D. Psillos and H. Niedderer, *Teaching and Learning in the Science Laboratory* (Kluwer, Dordrecht, 2002).
[92] L. J. Rennie, E. Feher, L. D. Dierking, and J. H. Falk, Toward an agenda for advancing research on science learning in out-of-school settings, J. Res. Sci. Teach. **40**, 112 (2003).
[93] C. A. Kubota and R. G. Olstadt, Effects of novelty-reducing preparation on exploratory behavior and cognitive learning in a science museum setting, J. Res. Sci. Teach. **28**, 225 (1991).
[94] D. Anderson and K. B. Lucas, The effectiveness of orienting students to the physical features of a science museum prior to visitation, Res. Sci. Educ. **27**, 485 (1997).
[95] N. Orion, A model for the development and implementation of field trips as an integral part of the science curriculum, School Sci. Math. **93**, 325 (1993).
[96] J. Huwer, *Nachhaltigkeit und Chemie im Schülerlabor: forschendes Experimentieren im Kontext einer naturwissenschaftlich-technischen Umweltbildung* (Universität des Saarlandes, Saarbrücken, 2015).
[97] M. Streller, The educational effects of pre and post-work in out-of-school laboratories, Ph.D. thesis, Technical University, Dresden, 2015.
[98] S. Osterlind, *Constructing Test Items* (Springer, Dortmund, 1997).

[99] T. Kubiszyn and G. D. Borich, *Educational Testing and Measurement: Classroom Application and Practice* (John Wiley, New York, 2013).
[100] X. Liu, Developing measurement instruments for science education research, *Second International Handbook of Science Education* (Springer, Dortrecht, 2012).
[101] A. Madsen, S. B. Mc Kagan, and E. C. Sayre, Resource letter RBAI-1: Research-based assessment instruments in physics and astronomy, Am. J. Phys. **85**, 245 (2017).
[102] H. Itzek-Greulich, B. Flunger, C. Vollmer, B. Nagengast, M. Rehm, and U. Trautwein, The impact of a science center outreach lab workshop on German 9th graders' achievement in science, in *Proceedings of the 10th Conference of the European Science Education Research Association* (ESERA, Nicosia, 2014), pp. 97–106.
[103] A. Helmke and F.-W. Schrader, School achievement, cognitive and motivational determinants, *International Encyclopedia of the Social and Behavioural Sciences* (Elsevier, Oxford, 2001), Vol. 20, pp. 13552–13556.
[104] L. Karam, Thematic issue: The interplay of physics and mathematics: Historical, philosophical and pedagogical considerations, Sci. Educ. **24**, 487 (2015).
[105] M. L. Fleming and M. R. Malone, The relationship of student characteristics and student performance in science as viewed by meta-analysis research, J. Res. Sci. Teach. **20**, 481 (1983).
[106] E. Wild, M. Hofer, and R. Pekrun, Psychologie des Lerners, *Pädagogische Psychologie—Ein Lehrbuch* (Beltz Psychologie Verlag, Weinheim, 2001), pp. 207–270.
[107] R. A. Louis and J. M. Mistele, The differences in scores and self-efficacy by student gender in mathematics and science, Int. J. Sci. Math. Educ. **10**, 1163 (2012).
[108] T. Mujtaba and M. J. Reiss, 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, Int. J. Sci. Educ. **35**, 1824 (2013).
[109] OECD, PISA 2006—Volume **2**: Data (OECD Publishing, Paris, 2007).
[110] J. Falk and L. Dierking, *The Museum Experience Revisited* (Left Coast Press, Walnut Creek, 2013).
[111] D. Hausamann, Extracurricular science labs for STEM talent support, Roeper Rev. **34**, 170 (2012).
[112] M. Boekaerts, Motivation, learning, and instruction, *The International Encyclopedia of the Social and Behavioral Science* (Elsevier, Oxford, 2001), pp. 10112–10117.
[113] P. R. Pintrich, Motivation and classroom learning, *Handbook of Psychology: Educational Psychology* (John Wiley, Hoboken, NJ, 2003), Vol. 7, pp. 103–122.
[114] H. L. Gibson and C. Chase, Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science, Sci. Educ. **86**, 693 (2002).
[115] A. Friedmann, *Framework for Evaluating Impacts of Informal Science Education Projekts* (National Science Foundation, Washington, DC, 2008), pp. 27–59.
[116] D. Sellmann and F. X. Bogner, Effects of a 1-day environmental education intervention on environmental attitudes and connectedness with nature, Eur. J. Psychol. Educ. **28**, 1077 (2013).
[117] J. W. Graham, Missing data analysis: Making it work in the real world, Ann. Rev. Psychol. **60**, 549 (2009).
[118] H. Itzek-Greulich, B. Flunger, C. Vollmer, B. Nagengast, M. Rehm, and U. Trautwein, Effectiveness of lab-work learning environments in and out of school: A cluster randomized study, Contemp. Educ. Psychol. **48**, 98 (2017).
[119] N. Garner and I. Eilks, The expectations of teachers and students who visit a non-formal student chemistry laboratory, Eurasia J. Math. Sci. Technol. Educ. **11** (2015).
[120] A. L. Luehman and D. Markowitz, Science teachers' perceived benefits of an out-of-school enrichment programme: Identity needs and university affordances, Int. J. Sci. Educ. **29**, 1133 (2007).
[121] I. Sasson and D. Cohen, Assessment for effective intervention: Enrichment science academic program, J. Sci. Educ. Technol. **22**, 718 (2013).
[122] A. A. Bartlett, The hydrostatic paradox revisited, Phys. Teach. **35**, 288 (1997).
[123] M. B. Partensky, Two boys and a can of coca cola, Phys. Teach. **40**, 106 (2002).
[124] J. Leisen, Das archimedische Prinzip—ein Lernzirkel mit Experimentier- und Problemlösesituationen, Raabits Phys. 1 (2005).
[125] S. Ganci, A multipurpose device for some hydrostatics questions, Phys. Teach. **46**, 407 (2008).
[126] R. De Luca and S. Ganci, A lot of good physics in the Cartesian diver, Phys. Educ. **46**, 528 (2011).
[127] N. Amir and R. Subramaniam, Making a fun Cartesian diver: A simple project to engage kineasthetic learners, Phys. Educ. **42**, 478 (2007).
[128] M. f. E. a. S. Rheinland-Pfalz, *Lehrplanentwurf Physik—Realschule, Gymnasium (Klassen 7/8–10)* (Sommer Druck und Verlag, Grünstadt, 1998).
[129] C. Bolte, Entwicklung und Einsatz von Erhebungsinstrumenten zur Analyse der Schüler-Lehrer-Interaktion im Chemieunterricht (Teil II), Z. Naturwissenschaften **3**, 67 (1997).
[130] C. Bolte and H. Stork, Entwicklung und Einsatz von Erhebungsinstrumenten zur Analyse der Schüler-Lehrer-Interaktion im Chemieunterricht (Teil I), Z. Naturwissenschaften **2**, 21 (1996).
[131] J. Kuhn, Authentische Aufgaben im theoretischen Rahmen von Instruktions-und Lehr-Lern-Forschung, *Optimierung von Ankermedien für eine neue Aufgabenkultur im Physikunterricht* (Wiesbaden, Vieweg + Teubner, 2010).
[132] J. Kuhn and A. Müller, Context-based science education by newspaper story problems: A study on motivation and learning effects, Prog. Sci. Educ. **2**, 5 (2014).
[133] A. Evers, C. Hagemeister, and A. Hostmaelingen, *EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests* (European Federation of Psychology Associations, Brussels, 2013).
[134] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, Phys. Rev. ST Phys. Educ. Res. **5**, 020103 (2009).
[135] D. G. Phillips, The development of the concept of displacement volume: A hierarchical model and its partial

testing under two methods of presentation, J. Res. Sci. Teach. **8**, 9 (1971).

[136] E. E. Glough and R. Driver, What do children understand about pressure in fluids?, Res. Sci. Technol. Educ. **3**, 133 (1985).

[137] P. A. Giese, Misconceptions about water pressure, in *Proceedings of the 2nd Int. Seminar, Misconceptions and Educational Strategies in Science and Mathematics* (Cornell University, Ithaca, NY, 1987), pp. 143–148.

[138] H.-C. She, Concepts of a higher hierarchical level require more dual situated learning events for conceptual change: A study of air pressure and buoyancy, Int. J. Sci. Educ. **24**, 981 (2002).

[139] N. Schmitt, Uses and abuses of coefficient alpha, Psychol. Assess. **8**, 350 (1996).

[140] K. S. Taber, The use of Cronbach's alpha when developing and reporting research instruments in science education, Res. Sci. Educ. **48**, 1273 (2018).

[141] W. K. Adams and C. E. Wieman, Development and validation of instruments to measure learning of expert-like thinking, Int. J. Sci. Educ. **33**, 1289 (2011).

[142] M. L. Van Blerkom, *Measurement and Statistics for Teachers* (Routledge, London, New York, 2009).

[143] M. D. Miller, R. L. Linn, and N. E. Gronlund, *Measurement and Assessment in Teaching* (Pearson, Upper Sadler River, NJ, 2009).

[144] D. Lang, C. Mengelkamp, and R. S. Jäger, Entwicklung von Testverfahren zur Berufsberatung von Schülern, Empirische Pädagogik **18**, 281 ff (2004).

[145] A. Kornmann and R. Horn, *SSB—Screeningverfahren für Schul- und Bildungsberatung* (Swets Test Services GmbH, Frankfurt a. M., 2001).

[146] A. Mayers, *Introduction to Statistics and SPSS in Psychology* (Pearson Education, Harlow, 2013).

[147] B. G. Tabachnik and L. S. Fidell, *Using Multivariate Statistics* (Pearson Education, Harlow, 2014).

[148] B. Wolf, *Handbuch pädagogische Psychologie* (Beltz, Weinheim, 2006).

[149] R. Hake, Interactive-engagement vs traditional methods. A six-thousend-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

[150] J. D. Marx and K. Cummings, Normalized change, Am. J. Phys. **75**, 87 (2007).

[151] J. von Korff, B. Archibeque, K. A. Gomez, T. Heckendorf, S. B. McKagan, E. C. Sayre, and L. Sorell, Secondary analysis of teaching methods in introductory physics: A 50 k-student study, Am. J. Phys. **84**, 969 (2016).

[152] P. Guderian, *Wirksamkeitsanalyse außerschulischer Lernorte—Der Einfluss mehrmaliger Besuche eines Schülerlabors auf die Entwicklung des Interesses an Physik* (Humboldt-Universität, Berlin, 2007).

[153] J. Wellington and J. Osborne, *Language and Literacy in Science Education* (Open University Press, Buckingham, UK, 2001).

[154] U. Trautwein, O. Lüdtke, B. W. Roberts I. Schnyder, and A. Niggli, Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement, J. Personality Social Psychol. **97**, 1115 (2009).

[155] N. Orion and A. Hofstein, Factors which influence learning ability during a scientific field trip in a natural environment, in *Proceedings of the Annual Convention of the National Association for Research in Science Teaching* (NARST, Fontana (IL), 1991).

[156] Nature Editorial, Learning in the wild, Nature (London) **464**, 813 (2010).

[157] S. von Stumm, B. Hell, and T. Chamorro-Premuzic, The hungry mind intellectual curiosity is the third pillar of academic performance, Perspect. Psychol. Sci. **6**, 574 (2011).

[158] F. Oppenheimer, The Exploratorium and Other Ways of Teaching Physics, Phys. Today **28**, 9, 9 (1975).

[159] J. H. Falk and L. D. Dierking, *Learning from Museums Visitor Experiences and the Making of Meaning* (AltaMira Press, Walnut Creek, CA, 2000).

[160] M. A. Skydsgaard, H. Möller-Andersen, and H. King, Designing museum exhibits that facilitate visitor reflection and discussion, Museum Management Curatorship **31**, 48 (2016).

[161] T. Jarvis and A. Pell, Factors influencing elementary school children's attitudes to science before during and following a visit to the UK National Space Centre, J. Res. Sci. Teach. **42**, 53 (2005).

[162] S. Asmussen, Lernen im Science-Center am Beispiel der Phänomenta (2010), http://www.widerstreit-sachunterricht.de.

[163] H. Hein, *The Exploratorium—The Museum as Laboratory* (Smithsonian Institution Press, Washington, DC, 1990).

[164] Technorama (2020), https://www.technorama.ch/en/scout/workshops-and-openlabs.

[165] R. Tytler, D. Symington, and J. C. Clark, Community-school collaborations in science: Towards improved outcomes through better understanding of boundary issues, Int. J. Sci. Math. Educ. **15**, 643 (2017).

[166] G. M. Koehl, Archimedes' principle and the hydrostatic paradox—simple demonstrations, Am. J. Phys. **17**, 579 (1949).

[167] E. Mullet and A. Montcouquiol, Archimedes' effect, information integration and individual differences, Int. J. Sci. Educ. **10**, 285 (1988).

[168] M. E. Loverude, C. H. Kautz, and P. R. Heron, Helping students develop an understanding of Archimedes' principle. I. Research on student understanding, Am. J. Phys. **71**, 1178 (2003).

[169] P. Heron, M. Loverude, P. Shaffer, and L. Mc Dermott, Helping students develop an understanding of Archimedes' principle. II. Development of research-based instructional materials., Am. J. Phys. **71**, 1188 (2003).

[170] P. Hewitt, *Conceptual Physics*, 12th ed. (Pearson/Addison-Wesley, San Francisco, 2014).