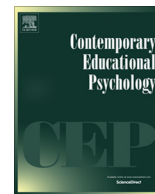




Contents lists available at ScienceDirect

## Contemporary Educational Psychology

journal homepage: [www.elsevier.com/locate/cedpsych](http://www.elsevier.com/locate/cedpsych)

## Empirical study

Effectiveness of lab-work learning environments in and out of school:  
A cluster randomized studyHeike Itzek-Greulich<sup>a,\*</sup>, Barbara Flunger<sup>a</sup>, Christian Vollmer<sup>b</sup>, Benjamin Nagengast<sup>a</sup>, Markus Rehm<sup>c</sup>,  
Ulrich Trautwein<sup>a</sup><sup>a</sup> Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Germany<sup>b</sup> Institute of Psychology, University of Education Heidelberg, Germany<sup>c</sup> Institute of Science, Technology & Geography, University of Education Heidelberg, Germany

## ARTICLE INFO

## Article history:

Available online 30 September 2016

## Keywords:

Cluster randomized field study

Science Center Outreach Lab (SCOL)

Achievement emotions

Student learning

State and trait motivation

## ABSTRACT

The issue of how to increase student motivation and achievement in science subjects is considered to be a major challenge in modern school systems. Lab-work learning environments in which students get direct (“hands-on”) experience with science content that is related to their everyday lives are posited to have positive effects on state/trait motivation and achievement, but there is a lack of sound empirical evidence to support this claim. In the present study, the effectiveness of a lab-work learning unit on the topic of “the chemistry of starch” was examined by applying a cluster randomized field study with three treatment conditions with lab-work elements and a control group. The first group was taught with lab-work elements in *School only*, the second group (*SCOL & school*) was taught in a combined condition encompassing both a SCOL (Science Center Outreach Lab) visit and classroom learning, the third group was taught entirely outside the school environment (*SCOL only*), and the fourth group was a wait-list control group, which was not exposed to a “starch” curriculum at the time of this study. Data from 1854 students were gathered in 67 ninth-grade classes on state motivation during the intervention and on trait motivation and achievement at pretest, posttest, and follow-up. Multilevel regression analyses revealed several differences between the lab-work conditions and the control group: Whereas the hands-on practical approach effectively enhanced state motivation with positive effects on joy, situational interest, situational competence, and reduced boredom in all three treatment conditions (*School only*, *SCOL & school*, and *SCOL only*), there were differences in trait effects: learning at school (*School only* and *SCOL & school*) increased achievement (posttest and follow-up), whereas the SCOL visit resulted in a small and spurious increase in trait motivation (reduced cost and increased competence beliefs only on the posttest).

© 2016 Elsevier Inc. All rights reserved.

## 1. Introduction

Students’ achievement motivation for science in school (e.g., their emotions, task values, interests, and perceived competence) tend to drop over the school years (Vierhaus, Lohaus, & Wild, 2016; Wigfield & Eccles, 2000; Zhu & Chen, 2010), and this drop may be especially pronounced in science subjects (Gottfried, Fleming, & Gottfried, 2001; Hofstein, Eilks, & Bybee, 2010; Todt & Schreiber, 1998). Motivation for science is the strongest predictor of achievement-related choices and performance (Wigfield, Tonks, & Klauda, 2009) in science-related college majors and careers (Chow, Eccles, & Salmela-Aro, 2012; Lau & Roeser, 2002). Accordingly, only a moderate number of students aim toward a

career in a science-related field (Butz, 2004; Committee on Prospering in the Global Economy of the 21st Century, 2007; European Commission, 2010; Osborne & Dillon, 2008; Schreiner & Sjøberg, 2004). At the same time, research has also shown that students can become interested in science classes when the topic at hand is directly related to their everyday life experiences (Stuckey, Hofstein, Mamlok-Naaman, & Eilks, 2013), is authentic (Adams, Gupta, & DeFelice, 2012), and includes many practical work phases (Hoffmann, 2002; Pugh, 2011; Stipek, 1996; Tytler, 2007). Application-oriented lessons can allow students to explore questions about their personal life contexts and can promote topic-related interest (Hofstein, Kesner, & Ben-Zvi, 1999; King & Ritchie, 2012; Sjøberg & Schreiner, 2006). However, in school, science teaching might not provide students with enough occasions to apply their knowledge, to do research, and to experiment independently (Hofstein et al., 2010).

\* Corresponding author.

E-mail address: [heikeitzek@gmail.com](mailto:heikeitzek@gmail.com) (H. Itzek-Greulich).

As a way to protect students' motivation from this negative trend and with the goal of adding to the typical learning that occurs in school or to enrich learning experiences in some way or another (Braund & Reiss, 2006; Klieme et al., 2010, p. 195), Science Center Outreach Labs (SCOLs) have been established in recent decades in many countries around the world, including Australia, Finland, Germany, Ireland, Israel, New Zealand, and the US (see Bolstad, Bull, Carson, Gilbert, & MacIntyre, 2013; Gomes & McCauley, 2012; Tytler, Osborne, Williams, Tytler, & Cripps Clark, 2008). SCOLs are quite expansive learning environments that are attended by an enormous number of students every year. In recent years, a number of studies (Ateşkan & Lane, 2016; Behrendt & Franklin, 2014; Fallik, Rosenfeld, & Eylon, 2013; Jensen & Lister, 2016; Patrick, Mathews, & Tunnicliffe, 2013; Rennie, 2014; Thomas, 2012) have investigated the effects of SCOLs. Unfortunately, however, many of these studies have provided only limited evidence for or against the effectiveness of SCOLs and have made only modest contributions to psychological theorizing for three reasons. First, so far, studies on the effects of SCOL visits have rarely used a randomized controlled experimental design (for an overview, see Priemer & Pawek, 2014), the lack of which has undermined the robustness of their findings against alternative explanations. We therefore used a cluster randomized field study to probe for the effects of a SCOL visit. Second, most studies have used a somewhat restricted set of motivational outcome variables that may not reflect the various dimensions of motivation that are involved in or result from a learning activity (e.g., Glowinski & Bayrhuber, 2011). Third, such studies have not systematically differentiated between the effects of including lab-work in regular school lessons and lab-work in out-of-school settings such as a SCOL.

Clearly, more studies are needed to target a broader range of motivational outcomes and to focus on students' motivational states during the SCOL visit as well as on their general motivation. Accordingly, relying on expectancy-value approaches (e.g., Eccles, 1983) and on interest theory (e.g., Krapp, 2000), the effects of lab-work learning environments on student motivation were investigated in the current study both during (using state measures) and before/after the respective lessons (using trait measures). The present study also extended research in this area by using data from an extensive project on the effectiveness of lab-work environments (see Itzek-Greulich et al., 2015) in which four distinct groups (*School only*, *SCOL & school*, *SCOL only*, and a control group) were compared. Previous analyses have shown that all three conditions featuring lab-work were effective at promoting students' achievement, irrespective of whether students visited a SCOL or not. The present study extended this previous work by (a) examining the effects of the same conditions on students' state and trait motivation, in particular students' emotional experiences while taking part in the different learning environments and carrying out experiments, and (b) exploring whether the previously identified effects on students' achievement could be found at follow-up.

### 1.1. Effectiveness of out-of-school learning

Out-of-school learning (e.g., visits to museums, science labs, zoos: Priemer, 2014; Schwan, Grajal, & Lewalter, 2014) is believed to result in immediate emotional responses and the motivation to get involved in a topic, and yet it might also affect students' general science-specific motivation. In fact, previous research has suggested that out-of-school visits have the potential to increase interest in science (amusement park: Singh, 2015; science center: Dairianathan & Subramaniam, 2011; Jarvis & Pell, 2005; university: Gibson & Chase, 2002; zoo visit/apes: Seybold, Braunbeck, & Randler, 2014), metacognitive engagement (amusement

park/physics: Nielsen, Nashon, & Anderson, 2009), and reasoning (natural history museum: Tenenbaum, To, Wormald, & Pegram, 2015). They have also been found to offer considerable emotional benefits such as increased joy, excitement (science centers, museums, and zoos: Rennie & McClafferty, 1995; university: Pluth, Boettcher, Nazin, Greenaway, & Hartle, 2015), and well-being; decreased anger, anxiety, and boredom (outdoor/amphibians: Randler, Ilg, & Kern, 2005); and increased choice (zoo visit/reptiles: Wünschmann, Wüst-Ackermann, Randler, Vollmer, & Itzek-Greulich, 2016).

However, the majority of studies that have investigated the effects of out-of-school learning (Hausamann, 2012; Luehmann, 2009; Meissner & Bogner, 2011; Thomas, 2012) on students' learning performance and motivation have been characterized by some limitations. Previous studies have relied on comparatively small sample sizes ( $N < 400$ ), with the exception of Seybold et al. (2014), and have not used a randomized design implemented in a multilevel context. Some of these studies (Dohn, 2011; Glowinski & Bayrhuber, 2011; Jarvis & Pell, 2005; Luehmann, 2009) were qualitative and did not include a control group. Finally, not many studies have compared out-of-school lab-work with lab-work learning environments from a regular school setting. Such studies are needed to disentangle effects of the learning environment (school vs. non-school) from the instructional approach (lab-work vs. no lab-work). In the following, we focus on SCOLs as a particular out-of-school learning facility.

### 1.2. SCOLs as a learning environment

It is often argued that the teaching of science in school does not provide students with enough occasions for "hands-on" contact with science (e.g., Abrahams & Reiss, 2012; Blanchard et al., 2010; Hart, Mulhall, Berry, Loughran, & Gunstone, 2000; Hofstein, Navon, Kipnis, & Mamlok-Naaman, 2005), although "hands-on" activities seem to evoke more interest in students (Swarat, Ortony, & Revelle, 2012). For this reason, some forms of lab-work education are now being used by many schools, although the ways in which they can implement this didactic approach are often limited. Like other out-of-school learning facilities (e.g., field trips, science/technology centers, museums, zoos, aquaria; for reviews, see DeWitt & Storksdiel, 2008; Rickinson et al., 2004), SCOLs offer exploration, discovery, and first-hand and original experiences. SCOLs come in many different forms in terms of ownership (private or public), content that is taught, and didactic approaches. However, there are a number of characteristics that are essential to (almost) all SCOLs. Most important, in contrast to school learning, SCOLs are closer to science application contexts, show a greater amount of authentic learning, and rely more on students engaging in practical work (Scharfenberg, Bogner, & Klautke, 2007).

In general, SCOLs resemble a workshop that is organized around structured lessons, and students are instructed by scientists who have the knowledge to help enhance the students' understanding of the natural sciences (Hausamann, 2012). In addition, to promote cognitive, affective, and psychomotor learning and to increase interest in science (Guderian, Priemer, & Schön, 2006), SCOLs also offer students the opportunity to get involved in experimental "hands-on" activities conducted by the students themselves (McClafferty & Rennie, 1993; Schwichow, Zimmerman, Croker, & Härtig, 2016). In typical SCOLs, students are given theoretical input on a certain topic and subsequently conduct experiments.

Students typically spend a couple of hours or a whole day in a SCOL; given the organizational constraints, longer units or repeated visits to SCOLs are the exception. For this reason, it has been argued (e.g., Glowinski & Bayrhuber, 2011; Itzek-Greulich et al., 2014; Scharfenberg et al., 2007) that SCOL visits may be

especially effective in terms of both short-term and long-term motivational outcomes, whereas their effects on the acquisition of structured knowledge and competencies may be less pronounced. Thus, the effectiveness of a SCOL with its emphasis on practical work may be linked to its ability to create a positive emotional setting, to spark enthusiasm, to foster students' domain-specific competence beliefs, and to positively influence students' interest in science subjects (Dohn, 2011).

### 1.3. Motivational outcomes of lab-work environments

The present study used expectancy-value theory (EVT; Eccles, 1983) and the person-object theory of interest (Krapp, Hidi, & Renninger, 1992) as its theoretical bases for exploring motivational effects.

Expectancy-value theory represents one of the most comprehensive theoretical models of students' motivation, specifying that students hold two kinds of motivational beliefs with regard to achievement-related settings: expectancy and value beliefs (Eccles, Wigfield, & Schiefele, 1998). Expectancy beliefs (Eccles et al., 1983) are understood as the subjective competence beliefs that one is able to accomplish academic tasks. Task values define people's reasons for engaging in a learning behavior (Eccles et al., 1983). Four major components of task values have been proposed: intrinsic value, attainment value, utility value, and cost (for a more detailed discussion of these components, see Eccles, 2005; Eccles & Wigfield, 2002; Gaspard et al., 2015). Intrinsic value is defined as the enjoyment gained from performing a task, utility value describes how useful the task is for future plans, attainment value is defined as the importance of doing well on a given task, and cost is the effort needed to engage in one activity and its negative impact on other valued activities (Eccles et al., 1983). Students who consistently exhibit (trait) motivation to develop science competencies (mastery goal) tend to be more engaged in science learning in and out of school than students with a high performance approach (Vedder-Weiss & Fortus, 2013).

With reference to expectancy-value approaches, one would expect effects of a SCOL visit on both students' general science-specific motivational beliefs ("traits" as a comparably stable personality characteristic) and students' immediate activity-related reactions ("states," which are sensitive to classroom contexts; see Pekrun, 2006; Pintrich, 2000). On the basis of EVT models, (a) the controllability of an activity (e.g., experienced through competence perceptions) and (b) the subjective value of the activity are proposed to be influential during achievement activities (states; e.g., Pekrun, 2006; Pekrun, Goetz, Titz, & Perry, 2002). Thereby, the learner's perceived control and value are claimed to be action-specific rather than science- or outcome-specific (e.g., Pekrun, 2006). Action-specific appraisals of control and value are defined to emerge as activity-related emotions (Pekrun, 2006). Activity-related emotions are emotions felt during work and learning (Pekrun, Elliot, & Maier, 2006); these can have a positive valence (e.g., joy) or a negative valence (e.g., anger or boredom). With regard to states that occur during the activities, therefore, students' activity-related emotions should be targeted, and in order to also directly capture activity-specific competences, their perceived competence during the out-of-school visit should be targeted.

As an additional theoretical foundation, the person-object theory of interest (Krapp et al., 1992) was used to obtain a broader understanding of students' situation- and person-specific reactions to lab-work learning environments. The person-object theory of interest distinguishes between students' situational interest as a reaction to characteristics of the tasks and the context versus their evolved individual interest, which is a relatively stable affective-evaluative orientation toward a domain (e.g., Schiefele, 2009). Situational interest refers to a rather temporary feeling (state),

whereas dispositional (individual) interest tends to be a sustainable (trait-like) feature of persons (Krapp et al., 1992). Situational interest encompasses *catch and hold components* (Mitchell, 1993). The catch component is directly connected to the activity. Application-oriented lessons can allow students to explore questions that pertain to their own personal life contexts and collect their own experiences. Thereby, such lessons can revive situational interest in the topics that are being taught (Krapp & Prenzel, 2011).

Previous studies have suggested that out-of-school learning can be used to increase students' situational and dispositional interest in science topics: When confronted with the atmosphere of authentic first-hand research conducted by scientists in a laboratory, and when students were given the opportunity to use professional equipment, students' situational (Dairianathan & Subramaniam, 2011; Seybold et al., 2014) and dispositional interest (Brandt, Möller, & Kohse-Höinghaus, 2008; Dohn, 2011; Gibson & Chase, 2002; Jarvis & Pell, 2005) increased. However, in addition to some methodological shortcomings, these studies left open the possibility that it was not the out-of-school visit (e.g., at a SCOL) per se that affected motivational outcomes but rather the opportunity for students to actively work on science experiments.

### 1.4. Achievement in lab-work environments

Lab-work, especially at a science center, promises to be effective for knowledge acquisition because it can extend the learning that occurs in school (Schwan et al., 2014). There are a number of studies on the influence of prior knowledge on emotional and motivational outcomes. For example, in the PISA study, achievement in science was associated with science enjoyment (Ainley & Ainley, 2011). Thus, students with prior knowledge on the topic of starch chemistry should experience more positive emotional and motivational outcomes than students without prior knowledge on the topic of the intervention, and therefore, when investigating the achievement outcomes, prior knowledge should be targeted.

Concerning the influence of lab-work in and out of school on achievement, previous research found that out-of-school learning leads to higher achievement than learning in school (Seybold et al., 2014; Sturm & Bogner, 2010; Wünschmann et al., 2016). However, these studies addressed informal learning and learning at museums and zoos, whereas studies on SCOLs are rare and suffer from some methodological limitations.

To overcome some of the limitations of previous research, Itzek-Greulich et al. (2015) compared achievement outcomes across three settings with a sufficiently large randomized field trial: *School only*, *SCOL & school*, and *SCOL only*. Although the intervention was effective in all three treatment groups (higher achievement than in the control group), the results indicated that students in the classroom learning condition (*School only*) and in the combined setting (*SCOL & school*) learned more than the students in the *SCOL only* condition. However, Itzek-Greulich et al. (2015) did not investigate motivational variables.

### 1.5. The present research

The present study was aimed at addressing the shortcomings of previous research. First, as highlighted above, the designs of most previous studies in terms of sample size, characteristics of the control group, lack of randomization, and statistical analyses limited their ability to detect causal effects that were due to out-of-school learning. To add to the existing knowledge, the present study used data from a randomized field trial with a pretest–posttest–follow-up design that differentiated between three forms of lab-work learning environments: *School only*, *SCOL & school*, and



SCOL only (see Itzek-Greulich et al., 2015) to test the effectiveness of a teacher-guided SCOL visit. This allowed us to check for differential effects of in-school and out-of-school learning. More specifically, in the study from which the current study took its data, Itzek-Greulich et al. (2015) (a) randomly assigned a large number of classes (67 classes of 14-to-15-year-old students, Grade 9) to one of four experimental conditions, (b) implemented a design with similar timing and concise contents (eight lessons on “the chemistry of starch”) across the three treatment groups and a wait-list control group, and (c) used multilevel analyses to address the clustering of the data.

Whereas most previous studies on the emotional and motivational gains of out-of-school learning have indicated positive motivational effects such as increased situational interest in the activities that were presented during the intervention (Dairianathan & Subramaniam, 2011; Seybold et al., 2014), increased dispositional interest in science subjects (Brandt et al., 2008; Dohn, 2011; Gibson & Chase, 2002; Jarvis & Pell, 2005), and more positive state emotions (Randler et al., 2005), these studies did not assess subjective task values (Eccles, 1983), or more generally, they did not assess state and trait components of motivation. Especially in the field of science education, where momentary interest may develop into long-lasting motivation for STEM subjects and career choices, the distinction between the four value components of motivation is an important issue. In the present study, achievement emotions, situational interest, and situational competence beliefs (state) were assessed subsequent to the intervention. Task values (attainment, cost, intrinsic value, and utility), dispositional interest, and competence beliefs (trait) were assessed before and after (posttest: 1 week after; follow-up: 6 weeks after) the intervention to compare the effects of the three learning environments on student motivation. We formulated the following hypotheses and research questions.

1. *Effects of lab-work on state motivation.* First, given that lessons with active student participation (e.g., lab-work) were found to be associated with increased student interest, a positive effect of the lab-work conditions on state motivation was expected (*School only*, *SCOL & school*, and *SCOL only*) in comparison with the control group on all state measures (Hypothesis 1).
2. *Effects of lab-work on trait motivation.* We also probed for possible effects of lab-work on the motivational trait measures. There has been a lack of support for effects on trait motivation in prior studies. However, we hypothesized that the effects of lab-work environments on motivation would be stronger in the conditions that included a Science Center Outreach Lab component (*SCOL & school*, *SCOL only*) than in the *School only* condition (Hypothesis 2). The rationale behind this is that the students would be experiencing the environment of the SCOL for the first time and this should be more exciting and should therefore lead to larger gains in trait motivation (Behrendt & Franklin, 2014) than in the teaching at school.
3. *Effects of lab-work on achievement.* Building on our previous research (Itzek-Greulich et al., 2015), we hypothesized that the *School only* condition would yield more pronounced effects on achievement than the other conditions (Hypothesis 3). This is suggested by the rationale that teachers with more profound knowledge about their students' learning mechanisms can do a better job of pointing out the most relevant key pieces of information to their students. In addition, whereas Itzek-Greulich et al. (2015) used five achievement subtests, we used a composite measure to investigate the overall effect of the intervention on students' academic achievement. We hypothesized that this overall measure would show more pronounced differences between the treatment groups.

## 2. Method

### 2.1. Sample

The data were collected from 1854 students in 67 classes from 22 secondary schools (Realschule) in the German state of Baden-Württemberg between November, 2012, and July, 2013 (see Itzek-Greulich et al., 2015). All participants (50.3% girls, 49.7% boys; 301 [16.2%] no gender specified) attended Grade 9 (intermediate track of secondary school) and were 15.3 years of age on average ( $SD = 0.7$ ). Trait measures of achievement and motivation were assessed 1 week prior to the intervention (pretest), 1 week after the intervention (posttest), and 6 weeks after the intervention (follow-up) at school in the students' regular classrooms in the presence of both a researcher and a teacher. Motivational state measures were assessed two times during the intervention. Active written consent for study participation was obtained from the schools' principals and from parents. Student participation was voluntary. Moreover, all teachers voluntarily participated in the study and agreed to be randomly assigned to any of the experimental conditions.

### 2.2. Experimental design

A randomized design was implemented with three experimental groups (*School only*, *SCOL & school*, and *SCOL only*) and a control group (see Fig. 1). Due to the nature of this whole-class intervention in which entire classes were randomized to treatment conditions, 17 of the 67 participating secondary school classes were randomly assigned to the *School only* group, 18 classes to the *SCOL & school* group, 17 classes to the *SCOL only* group, and 15 classes to the control condition.<sup>1</sup> Students' achievement emotions and motivational beliefs during the intervention (state) as well as their achievement, subjective task values, dispositional interest, and competence beliefs (trait) were assessed with questionnaires (see Fig. 1).

A letter with information concerning the background and content of the intervention and the topics covered by the questionnaires was sent to the administration of potentially participating schools within a radius of 150 km from the SCOL “EXPERIMENTA.” In order to guarantee that teachers were able to implement the intervention as intended, all participating teachers received an informational leaflet on the contents and timing of the intervention before the participating classes were randomized into the treatment conditions. That is, all participating teachers were briefed on the contents and the structure of the teaching unit prior to the implementation. The teachers participated in an individual 4-h informational session on how to structure the lesson, guided by the first author, and a manual on the teaching unit (discussed in the informational session) was provided to each teacher for guidance during the intervention. The classes were supplied with experimental kits and a CD-ROM containing the workbook as well as digital instructional material for the teacher.

#### 2.2.1. Description of the intervention

The treatment in the three experimental groups was the teaching unit (eight lessons of 45 min) on “the chemistry of starch.” For these three groups, the lessons were structured into a theoretical part (four lessons) and a practical part (four lessons). The students

<sup>1</sup> The classes in the *SCOL & school* group were distributed across the morning and afternoon lessons at the SCOL. In the control group, one class did not participate in the pretest, the intervention, the posttest, or the follow-up because the teacher was absent due to illness, and field time was limited by the end of the school year. Therefore, this class had to be excluded from the study. Fewer classes filled out the state-measure questionnaires during the intervention and the trait-measure questionnaires at follow-up (see Fig. 1).

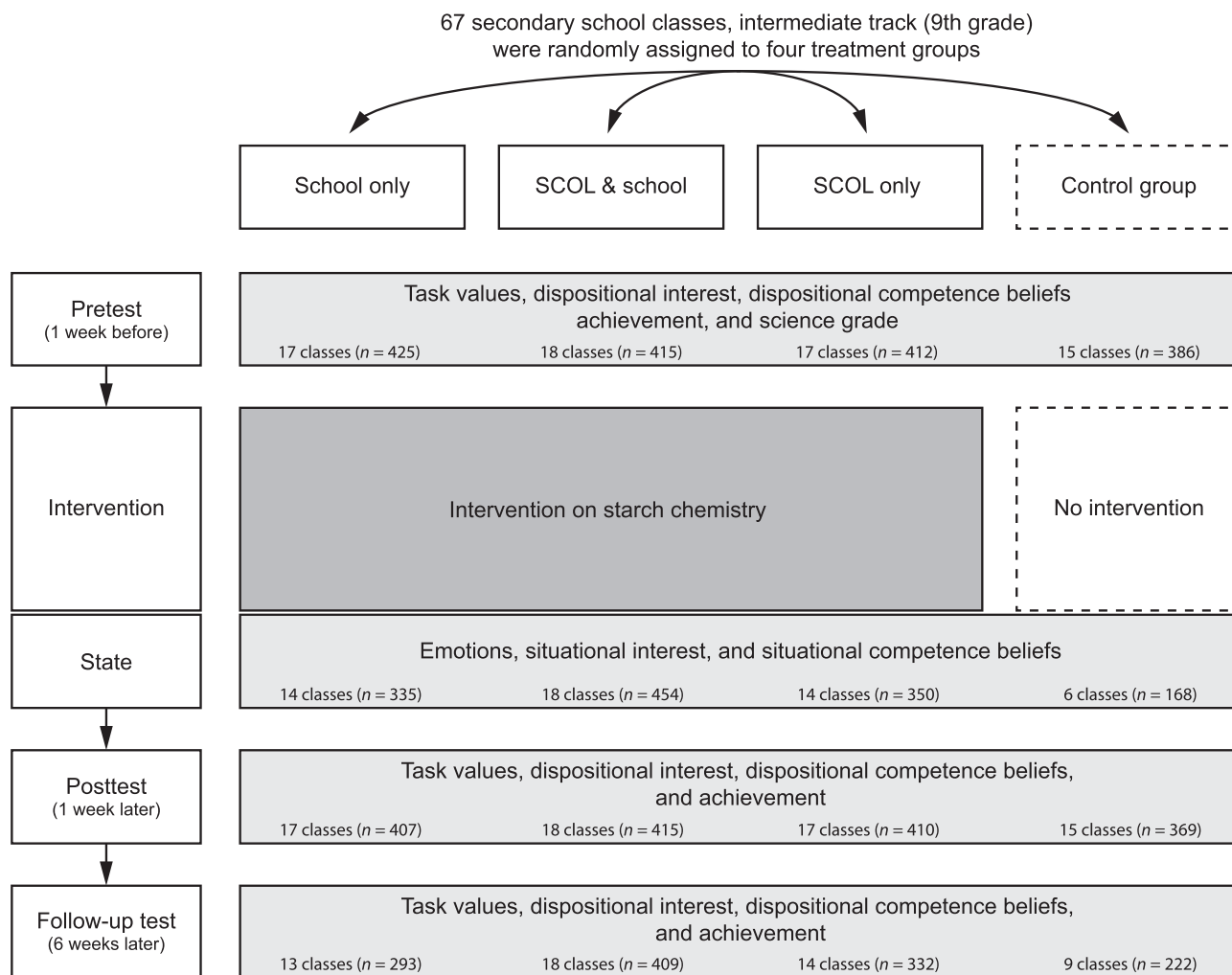


Fig. 1. Study design. Note. SCOL = Science Center Outreach Lab.

in the three treatment groups worked in small groups and got “hands-on” experience in science while conducting experiments on the topic of starch (extraction, microscope, production of films, glue, and compostable dishes).

The first group was taught in *School only* using mini-laboratory facilities that were provided to the schools and allowed the students to conduct most of the experiments that were also part of the SCOL visit. This *School only* group was taught the same topic by the regular science teacher. Teachers were supplied with materials similar to the ones used in the *SCOL only* condition. However, the scripts given to the *School only* group were designed by a teacher and had additional space for notes and introductions to experiments, structured like a workbook. The students filled out these workbooks. The first unit covered chemical tests of carbohydrates, whereas the second unit focused on the chemistry of starch, starch as a renewable raw material, and the application of starch (Itzek-Greulich et al., 2014). A lab in miniature form was provided, and a training session was given to teachers who were not familiar with this material. Teaching methods included open learning and experiments.

A foundation-operated interactive student research center in South Germany (“EXPERIMENTA”) was chosen for the *SCOL & school* and *SCOL only* conditions. The course was an established part of the regular curriculum taught at the SCOL and was not altered for our study. The SCOL provided their two most experienced trainers for the workshops; both trainers covered an equal number of

workshops. The workshops were taught by one scientist supported by a lab assistant who supplied and cleaned the materials. The SCOL scientists were not regular school teachers but had received some training in didactics, which is a typical and, in some sense, integral part of the SCOL concept. The scientist and assistant were assigned to classes by the SCOL management. In line with typical regulations, the usual science teacher from school supervised the students at the SCOL.

The second group (*SCOL & school*) was taught in a combined condition that encompassed both a SCOL visit and classroom learning (embedded). In this way, extracurricular learning experiences were integrated into the classroom lessons. The classes in the *SCOL & school* condition were given two introductory lessons at school by the regular science teacher, each 45 min long, with workbooks and material similar to the *School only* condition but shortened to meet the available time at school. The lessons involved a box of chemical tools and materials provided by the researcher. Subsequently, they visited the SCOL for half a day and were taught about the chemical topic of starch by the lab scientist and assistant as in the *SCOL only* condition. Afterwards, two school lessons were devoted to repetition and consolidation. The *SCOL & school* condition was given a combination of scripts and materials from both institutions. The total teaching time was the same as in the other two experimental conditions.

For the third group (*SCOL only*), the whole intervention was taught at the SCOL. The students in the *SCOL only* condition worked

with material and an original script from the SCOL; this script had a theoretical part followed by experimental instructions. Prior to the intervention, the first author met the two scientists, their assistants, and the pedagogical director at the SCOL two times to ensure that they were familiar with the intervention topics and state questionnaires applied in the SCOL course. The students worked in small groups. They got “hands-on” experience in science with a scientist from the SCOL and filled out scripts on the topic that were taken back to school by the students. The compact science workshop was not structured around the 45-min lessons that are typical of German schools but in longer units of 4 h. A specific focus of the science learning was practical work; only little prior theoretical knowledge was needed. The SCOL workshop basically consisted of two parts: First, the students were dressed in lab coats and received a detailed safety briefing before they began working, followed by an introduction to carbohydrate chemistry, including corresponding detection techniques. After their lunch break during which they left the lab, the second part followed (“starch”: extraction, microscope, production of films, glue, and compostable dishes). The day at the SCOL began with a short introduction and ended with a summary by the SCOL scientist.

After the *School only* and *SCOL & school* interventions, the experimental kits were retrieved and resupplied for the next class. This also provided nonreactive evidence for the fact that the intervention was implemented because the chemical contents of the experimental kits were used up by the students when the exercises were completed as intended.

#### 2.2.2. The control condition

The fourth learning group was a control group that was not exposed to the intervention (“starch” curriculum) at the time of this study. In these classes, the topic of the intervention was taught later in the school year or in the next school year. Thus, the control group was a wait-list control group, which offers the advantage that we were able to determine the “total” effect of the three intervention conditions as compared to a situation without treatment. The untreated control group received their regular chemistry lessons. The topics of these lessons (alcohols, acids, fossil fuels, alkane, and atoms) were not related to the study topic

#### 2.4. Measures

Regarding state and trait motivational measures, participants responded on a Likert scale that ranged from 1 (*disagree*) to 4 (*agree*).

##### 2.4.1. Achievement emotions, situational interest, and situational competence beliefs (state)

Achievement emotions, situational interest, and situational competence beliefs were studied as outcomes directly after the theoretical and practical lessons in the teaching unit. The learning-related emotion scale by Pekrun, Goetz, Frenzel, Barchfeld, and Perry (2011) was used to assess students' achievement emotions: anger (three items, e.g., “I got angry in class”), joy (three items, e.g., “I had fun in class”), and boredom (three items, e.g., “I was so bored that I stopped following the lesson”). Situational interest was measured with three items (e.g., “It was interesting”). Situational competence beliefs were measured with a three-item scale (e.g., “I followed the lessons easily”) developed by Willems (2011). For the state measures, Cronbach's  $\alpha$  ranged from 0.82 to 0.87. See Table 1 for correlations and descriptive statistics for state motivation during the practical part of the lesson. The correlations between the measures from the theoretical and practical parts were consistently high (from 0.61 to 0.68).

##### 2.4.2. Task values, dispositional interest, and competence beliefs (trait)

Trait motivation was measured three times (at pretest, posttest, and follow-up) with six scales assessing subjective task values, dispositional interest, and competence beliefs.

**2.4.2.1. Task values.** Subjective value beliefs regarding science were assessed for all four components of the EVT (Eccles et al., 1983): attainment, cost, intrinsic, and utility. To this end, Trautwein et al.'s (2012) items were adapted to the domain of science. Attainment value was measured with three items (e.g., “Science is important to me personally”;  $\alpha = 0.79$ ). Cost was measured with two items (e.g., “I'd have to sacrifice a lot of free time to be good at science”;  $\alpha = 0.83$ ). Intrinsic value was measured with four items (e.g., “I enjoy puzzling over science problems”;  $\alpha = 0.82$ ). Utility value was measured with three items (e.g., “I'll need good science skills for my later life [education, training, studies, work]”;  $\alpha = 0.74$ ).

**2.4.2.2. Dispositional interest.** Students' dispositional interest in science was assessed with three items (e.g., “I am interested in scientific topics”;  $\alpha = 0.84$ ; adapted from Pekrun, Götz, Zirngibl, & Jullien, 2002).

**2.4.2.3. Competence beliefs.** Students' general science competence beliefs were measured with three items from the German adaptation (Schwanzer, Trautwein, Lüdtke, & Sydow, 2005) of the Self-Description Questionnaire III (Marsh, 1992), a multidimensional competence beliefs instrument for older adolescents and young adults (sample item: “I have always been good at science;  $\alpha = 0.87$ ).

##### 2.4.3. Achievement

Overall achievement was measured with a multiple-choice knowledge test (Itzek-Greulich et al., 2015) that consisted of four subtests (Carbohydrate Specific Knowledge, Chemical Analysis, Experimental Specific Knowledge, and Declarative Knowledge). An additional achievement subtest (Chemical Terms) relied on self-reports of the participants' familiarity with chemical terms, and therefore this subtest was not included in the present study's analyses. The knowledge test covered a broad range of topics on starch chemistry as well as on carbohydrates, including content-specific knowledge on the topics lab-work, starch, photosynthesis, carbohydrates, declarative knowledge of organic chemistry, and practical skills needed for a Fehling test (see Appendix A and Itzek-Greulich et al., 2015). The four subtests (Carbohydrate Specific Knowledge, Chemical Analysis, Experimental Specific Knowledge, and Declarative Knowledge) were integrated into an overall measure of achievement (45 items for the posttest and follow-up), whereas the pretest contained 33 items (Carbohydrate Specific Knowledge was not tested on the pretest). Example item: “Which is the main task of chloroplasts in plants? (a) Absorbing light energy and producing nutrients, (b) removing metabolic waste by active transport, (c) producing chemical energy from nourishment, (d) controlling the form of cells.” Achievement scores were computed with item response theory in the R package “mIRT” (Chalmers, 2012). EAP factor scores were estimated with the unidimensional 2PL IRT model by applying item response theory (pretest: EAP reliability = 0.66, RMSEA = 0.10; posttest: EAP reliability = 0.81, RMSEA = 0.08, follow-up: EAP reliability = 0.72, RMSEA = 0.07). The factor scores were used for subsequent analyses.

##### 2.4.4. Science grade

Science grade was measured as a control variable. Students were asked to report the grade from their last school certificate (self-report). The original scale is given on an integer scale that

**Table 1**

Correlations and descriptives for emotions, situational interest, and situational competence beliefs (state, practical part).

		1	2	3	4	M	SD	N	Missings (%)		ICC
									Student level	Class level	
1	Anger					1.81	0.73	1216	13.4	22.4	0.13
2	Joy	−0.61				2.74	0.75	1217	13.3	22.4	0.19
3	Boredom	0.71	−0.67			1.79	0.77	1215	13.5	22.4	0.14
4	Situational interest	−0.55	0.79	−0.66		2.68	0.69	1219	13.2	22.4	0.14
5	Situational competence	−0.41	0.587	−0.45	0.54	3.10	0.63	1218	13.2	22.4	0.15

Note. Pearson's correlations. All correlations were highly significant ( $p < 0.001$ ). Scale from 1 = "disagree" to 4 = "agree". ICC = intraclass correlation.

ranges from 6 (failing) to 1 (outstanding) and was recoded to 1 (failing) to 6 (outstanding).

## 2.5. Statistical analyses

### 2.5.1. Confirmatory factor analyses

Tests of measurement invariance (Meredith, 1993; Widaman & Reise, 1997) were conducted to assess the structural stability (validity) of trait motivation and achievement over time in Mplus 7 (Muthén & Muthén, 1998–2012). More precisely, measurement invariance was tested by comparing three nested models (reflecting configural, metric, and scalar invariance) with increasing invariance constraints (e.g., Meredith, 1993) separately for each dependent variable. The first model tested the invariance of the factor structure (i.e., configural invariance). The second model entailed constrained factor loadings over time (i.e., metric invariance). The third model imposed invariant factor loadings and invariant item intercepts over time, thus testing for scalar measurement invariance. The models were evaluated by following suggestions made by Chen (2007) and Cheung and Rensvold (2002). That is, when the decrease in fit for the more restrictive model was less than 0.01 for the incremental fit indices (e.g., CFI), and when the RMSEA differed by less than 0.015, the more restrictive model was preferred. The comparisons across time suggested a good fit regarding metric invariance (see Appendix B) and supported the increases in the invariance constraints on all trait measures.

### 2.5.2. Multilevel regression analyses

After measurement invariance was established, we computed multilevel regression analyses to test our predictions for each dependent variable with three class-level dummy variables that indicated the lab-work conditions (*School only*, *SCOL & school*, *SCOL only*) and with the control condition as the reference group. Differences between the lab-work groups were investigated by specifying equivalent alternative groups with different codes for the different experimental groups.

For the trait measures, the models included the initial level of the respective outcome on the pretest as a covariate at the student level and at the class level. The effects at both the student and class levels were freely estimated. At the student level, the covariate was group-mean centered (Enders & Tofghi, 2007). Manifest aggregation was used for the class-level predictor (Marsh et al., 2009). At the class level, the covariate was grand-mean centered (Enders & Tofghi, 2007). All variables except for the indicator variables were standardized prior to the analyses.

### 2.5.3. Missing data

Due to the absence of students at single measurement points and nonresponses to single items, data were missing at rates of 13–14% for the state measures (Table 1) and at 12–16% for the trait

measures on the student level (see Table 2). The amount of missing data on the student level was highest at the follow-up (about 16% for all variables). The levels of missing data on the class level were relatively high for the state measures and at the follow-up time point. The state questionnaires were handed out to the teachers prior to the intervention, and the teachers were asked to administer the state questionnaires. However, some of the teachers did not give the state questionnaires to the students, and therefore, a considerable amount of missing data occurred on the class level. Concerning these data that were missing at the class level, in the control group, the classes with missing data did not differ significantly from the other classes in achievement or state motivation at pretest. However, there were significant differences at pretest concerning the classes from the *School only* (three classes) and *SCOL only* (three classes) conditions that did not fill out the state measures: These six classes had lower values on achievement (both conditions), attainment (*School only*), intrinsic value (*School only*), utility value (*SCOL only*), dispositional interest (both), and competence beliefs (*SCOL only*). Nevertheless, the posttest results did not change when the classes that did not fill out the state measures (3 classes in the *School only*, 3 classes in the *SCOL only*, and 11 classes in the control group) were not included in the posttest analyses. For attrition on the class level for the state measures and at follow-up, see also Fig. 1 and Tables 1 and 2. Full information maximum likelihood (FIML; Arbuckle, 1996; Enders, 2010) was used to take the missing data into account. Pretest scores on trait motivation, achievement, and grades were included as predictors in the multilevel regression analyses to make the assumption of missing-at-random more plausible.

## 3. Results

### 3.1. Preliminary analyses

In the first step, a randomization check was performed to test for pre-intervention differences in the motivational and achievement measures across the four groups (three treatment groups and a control group) via pairwise comparisons of means in a multilevel regression model. We found no significant differences between the groups in the motivational trait measures at pretest. Concerning the achievement measures, there was one significant difference: the *SCOL only* group had significantly higher scores than the control condition ( $p = 0.009$ , see Fig. 3), but no other comparison was statistically significant. Nevertheless, we included the pre-intervention scores to increase the power for testing treatment effects and to control for any remaining differences between the groups.

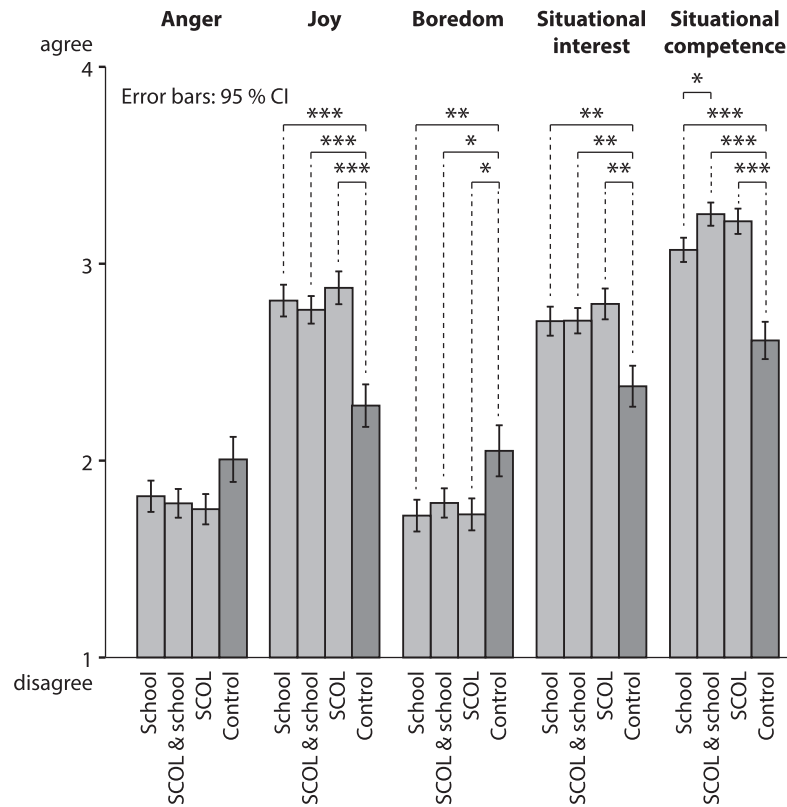
The intraclass correlations (ICCs; Tables 1 and 2) were higher for the motivational state variables and for achievement than for the motivational trait variables. Concerning achievement, the variation between treatments (ICC) was higher after the treatment.

**Table 2**  
Correlations and descriptives for achievement, task values, dispositional interest, and competence beliefs (trait).

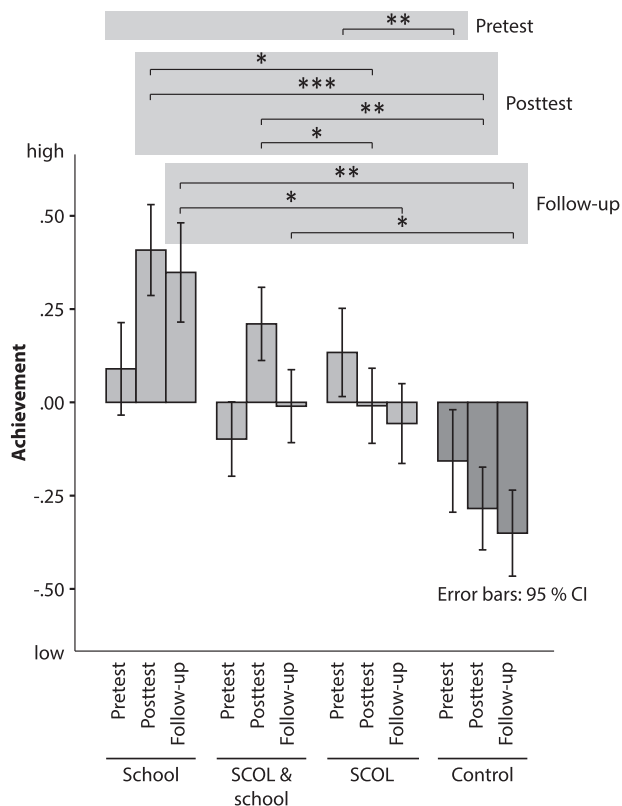
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	<i>M</i>	<i>SD</i>	<i>N</i>	Missings (%)		ICC
																									Student level	Class level	
1	Achievement (pretest)																					0.00	1.00	1639	11.6	0.0	0.11
2	Achievement (posttest)	0.32																				0.00	1.00	1601	13.6	0.0	0.23
3	Achievement (follow-up)	0.37	0.49																			0.00	1.00	1257	15.2	19.4	0.25
4	Attainment value (pretest)	0.13	0.18	0.20																		2.21	0.64	1631	12.0	0.0	0.05
5	Cost (pretest)	−0.09	−0.12	−0.11	−0.27																	2.51	0.82	1635	11.8	0.0	0.03
6	Intrinsic value (pretest)	0.16	0.21	0.17	0.77	−0.33																2.29	0.63	1635	11.8	0.0	0.05
7	Utility value (pretest)	0.12	0.16	0.17	0.62	−0.15	0.61															2.31	0.69	1630	12.1	0.0	0.03
8	Dispositional interest (pretest)	0.17	0.19	0.18	0.72	−0.28	0.78	0.63														2.15	0.61	1637	11.7	0.0	0.03
9	Competence beliefs (pretest)	0.17	0.20	0.16	0.57	−0.57	0.66	0.42	0.61													2.57	0.61	1632	12.0	0.0	0.03
10	Attainment value (posttest)	0.13	0.17	0.14	0.62	−0.25	0.61	0.46	0.60	0.47												2.24	0.65	1594	14.0	0.0	0.06
11	Cost (posttest)	−0.10	−0.15	−0.09	−0.25	0.50	−0.31	−0.18	−0.26	−0.50	−0.28											2.42	0.83	1600	13.7	0.0	0.04
12	Intrinsic value (posttest)	0.15	0.21	0.17	0.58	−0.26	0.66	0.45	0.62	0.52	0.79	−0.32										2.29	0.63	1600	13.7	0.0	0.04
13	Utility value (posttest)	0.15	0.16	0.17	0.52	−0.18	0.50	0.56	0.53	0.38	0.65	−0.18	0.63									2.28	0.69	1595	14.0	0.0	0.02
14	Dispositional interest (posttest)	0.14	0.18	0.16	0.58	−0.24	0.62	0.49	0.67	0.49	0.75	−0.27	0.78	0.64								2.17	0.64	1598	13.8	0.0	0.04
15	Competence beliefs (posttest)	0.17	0.25	0.20	0.49	−0.47	0.56	0.36	0.52	0.69	0.60	−0.59	0.65	0.47	0.61							2.56	0.63	1595	14.0	0.0	0.05
16	Attainment value (follow-up)	0.12	0.14	0.11	0.59	−0.24	0.58	0.46	0.57	0.45	0.69	−0.29	0.64	0.52	0.61	0.51						2.24	0.66	1246	16.0	19.4	0.04
17	Cost (follow-up)	−0.10	−0.12	−0.10	−0.24	0.46	−0.28	−0.18	−0.25	−0.43	−0.27	0.53	−0.29	−0.16	−0.27	−0.49	−0.29					2.45	0.85	1251	15.6	19.4	0.02
18	Intrinsic value (follow-up)	0.10	0.16	0.13	0.56	−0.26	0.63	0.46	0.58	0.48	0.66	−0.32	0.71	0.50	0.62	0.55	0.78	−0.32				2.26	0.63	1252	15.6	19.4	0.05
19	Utility value (follow-up)	0.12	0.15	0.13	0.48	−0.19	0.49	0.56	0.54	0.36	0.56	−0.23	0.55	0.62	0.57	0.42	0.67	−0.19	0.63			2.23	0.71	1249	15.8	19.4	0.02
20	Dispositional interest (follow-up)	0.12	0.18	0.14	0.56	−0.25	0.60	0.50	0.63	0.47	0.67	−0.27	0.68	0.57	0.71	0.54	0.73	−0.25	0.76	0.68		2.13	0.64	1250	15.7	19.4	0.04
21	Competence beliefs (follow-up)	0.15	0.19	0.19	0.44	−0.45	0.50	0.32	0.47	0.61	0.49	−0.51	0.52	0.37	0.50	0.71	0.56	−0.58	0.60	0.45	0.56	2.50	0.63	1249	15.8	19.4	0.04

Note. Pearson's correlations. All correlations were highly significant ( $p < 0.001$ ). Trait motivation on a scale from 1 = "disagree" to 4 = "agree"; Achievement: IRT-scaled and standardized (mean = 0, SD = 1). ICC = intraclass correlation.





**Fig. 2.** Emotions, situational interest, and situational competence beliefs (state) by treatment. Note. The bar graph depicts raw scores and significant differences between the treatment groups shown in Table 3. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .



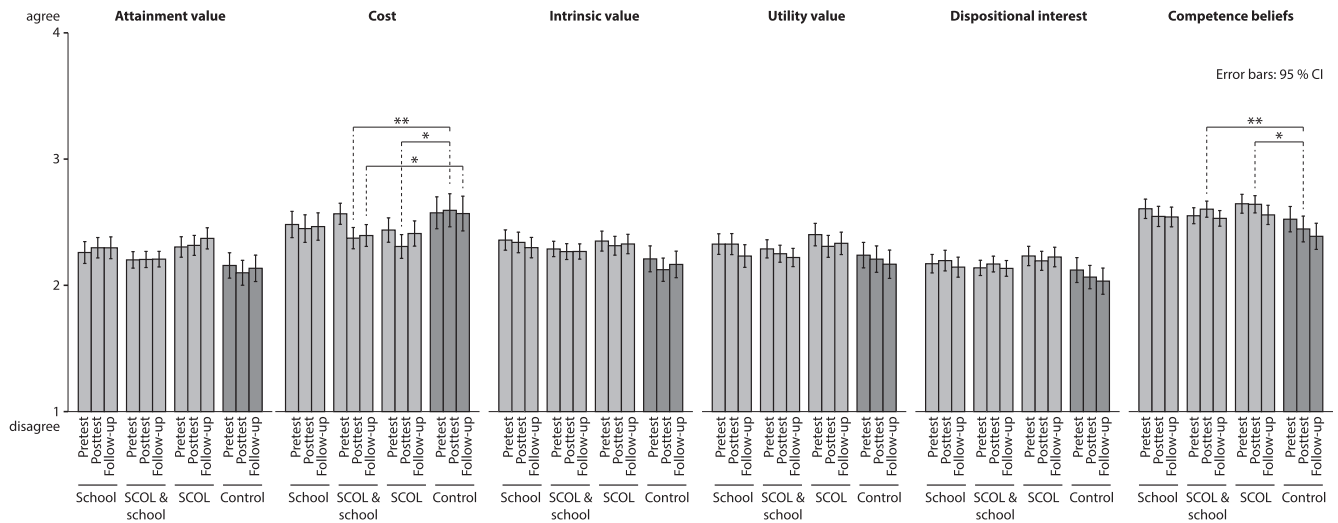
**Fig. 3.** Achievement by treatment and testing time. Note. The bar graph depicts IRT scores and significant differences between the groups from the multilevel analyses. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

### 3.2. Descriptives and bivariate correlations

The descriptive statistics (means and standard deviations) and the intercorrelations of all outcome variables are shown in Table 1 (state measures) and Table 2 (trait measures). Figs. 2–4 present the means of all the outcome variables by treatment group (see also Table 3). The students' responses were most positive for situational competence and highest for joy as well as situational interest, whereas mean values for anger and boredom were generally lower (Fig. 2). Corresponding to this and regarding trait motivation, students rated their competence beliefs the highest; however, the students also judged the cost of engaging in science topics relatively high (Fig. 4). Correlations between state measures were highest for boredom and anger as well as for situational interest and joy (Table 1). Achievement was weakly correlated with trait motivation. Concerning trait motivation, cost was negatively correlated with all other motivational trait variables; correlations among the trait motivation variables were generally of medium strength ( $r = 0.32$ – $0.78$ ), whereby cost and utility value had the lowest correlation coefficients ( $r = -0.15$ – $-0.18$ – $-0.19$ ) at pretest, posttest, and follow-up (Table 2).

### 3.3. Treatment effects on state motivation

The effects of the lab-work conditions (School only, SCOL & school, and SCOL only) during the practical part of the intervention on each state measure were tested via separate multilevel analyses for each outcome; the control group served as the reference group (Table 4, Fig. 2). Trait motivation, achievement, and science grade were included as covariates. In comparison with the control group, the students in the three intervention groups generally reported



**Fig. 4.** Task values, dispositional interest, and competence beliefs by treatment and testing time. *Note.* The bar graph depicts raw scores and significant differences between the treatment groups shown in Tables 4 and 5. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

**Table 3**  
Descriptives for the outcomes by treatment.

		School only			SCOL & school			SCOL only			Control		
		M	SD	N	M	SD	N	M	SD	N	M	SD	N
State (practical part)	Anger	1.82	0.70	304	1.78	0.74	403	1.75	0.72	341	2.00	0.75	168
	Joy	2.81	0.71	304	2.77	0.71	403	2.88	0.78	342	2.28	0.71	168
	Boredom	1.72	0.71	305	1.78	0.75	400	1.72	0.76	342	2.05	0.85	168
	Situational interest	2.70	0.65	306	2.70	0.65	403	2.79	0.73	342	2.38	0.68	168
	Situational competence	3.06	0.54	306	3.25	0.60	402	3.21	0.60	342	2.61	0.62	168
Posttest	Achievement	0.41	1.09	474	0.21	1.01	473	−0.01	0.94	475	−0.28	0.84	430
	Attainment	2.26	0.64	406	2.21	0.61	414	2.29	0.64	408	2.20	0.70	366
	Cost	2.46	0.85	407	2.37	0.81	415	2.35	0.80	410	2.54	0.86	368
	Intrinsic	2.33	0.65	407	2.27	0.60	415	2.31	0.61	410	2.22	0.66	368
	Utility	2.31	0.68	406	2.26	0.65	414	2.30	0.70	407	2.28	0.73	368
	Dispositional interest	2.20	0.66	406	2.18	0.60	415	2.19	0.63	410	2.12	0.66	367
	Competence beliefs	2.54	0.63	407	2.60	0.61	415	2.62	0.59	406	2.46	0.67	367
Follow-up	Achievement	0.35	1.10	292	−0.01	0.95	411	−0.06	0.95	332	−0.35	0.84	222
	Attainment	2.26	0.70	291	2.22	0.58	409	2.35	0.69	326	2.12	0.70	220
	Cost	2.48	0.86	292	2.39	0.81	409	2.42	0.83	330	2.55	0.93	220
	Intrinsic	2.26	0.66	293	2.27	0.56	409	2.32	0.64	330	2.15	0.70	220
	Utility	2.21	0.71	293	2.23	0.67	408	2.32	0.74	328	2.15	0.75	220
	Dispositional interest	2.11	0.64	292	2.14	0.59	408	2.22	0.66	330	2.02	0.70	220
	Competence beliefs	2.51	0.62	292	2.52	0.58	408	2.54	0.63	329	2.38	0.70	220

*Note.* State and trait motivation: Likert scale from 1 = “disagree” to 4 = “agree”; Achievement: IRT-scaled and standardized (mean = 0, SD = 1).

less boredom as well as more joy, situational interest, and situational competence. There were no significant differences on anger. Thus, in line with our first prediction, we found a positive effect of the lab-work conditions in comparison with the control group on the state measures except for anger.<sup>2</sup> The effects were most pronounced for situational competence. There were no significant differences between the three treatment groups with the exception of situational competence; that is, the three lab-work conditions were all effective at eliciting more state motivational gains than the

control group. In situational competence, the SCOL & school condition had higher values than the School only condition.

### 3.4. Treatment effects on trait motivation

The effects of the lab-work conditions on motivation were again tested via separate multilevel analyses for each outcome; the control group served as the reference group (see Table 4 for the posttest and Table 6 for the follow-up, Fig. 4). In each analysis, we controlled for the pretest value of the corresponding outcome at the student and class levels. Overall, two of the five trait outcomes were affected by the experimental conditions (cost and competence beliefs, Tables 4 and 5, Fig. 4). Students in the SCOL & school and in the SCOL only conditions rated their cost significantly lower than students in the control condition; in the SCOL & school condition, the effect still held after 6 weeks (Table 5). Students in the SCOL & school and the SCOL only conditions reported higher

<sup>2</sup> Overall, the practical part of the lab-work elicited more positive effects on state motivation than the theoretical part. When comparing the effects of the lab-work to the control group, the effects on joy and situational competence were most pronounced in both the theoretical and practical parts, whereas the effects of the lab-work on boredom and situational interest were less pronounced in the theoretical part of the lesson. There were no differences between the three lab-work conditions in state motivation during the theoretical part of the lesson.

**Table 4**

Effects of the lab-work conditions on students' motivational variables during the intervention (state, practical part).

	Anger			Joy			Boredom			Situational interest			Situational competence		
	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>
<i>Student level</i>															
Attainment value (pretest)	−0.10	(0.06)	†	0.06	(0.05)		−0.08	(0.06)		0.06	(0.06)		−0.02	(0.05)	
Cost (pretest)	0.02	(0.05)		0.03	(0.04)		0.02	(0.04)		0.00	(0.04)		−0.03	(0.04)	
Intrinsic value (pretest)	−0.04	(0.07)		0.13	(0.06)	*	−0.07	(0.07)		0.17	(0.06)	**	0.07	(0.05)	
Utility value (pretest)	0.02	(0.04)		−0.04	(0.03)		0.02	(0.03)		−0.02	(0.03)		0.00	(0.04)	
Dispositional interest (pretest)	−0.03	(0.04)		0.07	(0.05)		−0.04	(0.05)		0.11	(0.05)	*	0.03	(0.05)	
Competence beliefs (pretest)	−0.04	(0.06)		0.09	(0.05)	†	−0.08	(0.05)		0.04	(0.05)		0.20	(0.04)	***
Achievement (pretest)	−0.10	(0.04)	*	0.07	(0.03)	*	−0.09	(0.03)	**	0.04	(0.03)		0.07	(0.04)	†
Science grade	0.01	(0.03)		0.02	(0.03)		0.00	(0.03)		−0.01	(0.03)		−0.03	(0.04)	
<i>Class level</i>															
School only	−0.19	(0.13)		0.69	(0.15)	***	−0.36	(0.11)	**	0.41	(0.13)	**	0.63	(0.12)	***
SCOL & school	−0.28	(0.15)	†	0.70	(0.16)	***	−0.29	(0.13)	*	0.50	(0.15)	**	1.00	(0.12)	***
SCOL only	−0.20	(0.15)		0.73	(0.18)	***	−0.30	(0.13)	*	0.54	(0.15)	**	0.82	(0.11)	***
<i>R<sup>2</sup></i>															
Student level	0.06			0.09			0.08			0.13			0.11		
Class level	0.09			0.32			0.15			0.27			0.66		
Snijders & Bosker	0.12			0.17			0.12			0.17			0.20		
ICC	0.15			0.20			0.15			0.14			0.17		
N	1002			1004			1004			1005			1004		

Note. At the student level, the covariate attainment value was group-mean centered. All continuous variables were z-standardized beforehand. The reference group for the treatment was the control group.  $\beta$  = standardized regression coefficient. Snijders & Bosker  $R^2$ : mean squared prediction error following Snijders and Bosker (1994).

†  $p < 0.10$ .  
 \*  $p < 0.05$ .  
 \*\*  $p < 0.01$ .  
 \*\*\*  $p < 0.001$ .

**Table 5**

Effects of the lab-work conditions on students' learning achievement, subjective task values, dispositional interest, and competence beliefs (posttest).

	Achievement			Attainment value			Cost			Intrinsic value			Utility value			Dispositional interest			Competence beliefs		
	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>
<i>Student level</i>																					
Attainment value (pretest)	−0.03	(0.04)		0.57	(0.03)	***															
Cost (pretest)	0.00	(0.03)					0.44	(0.03)	***												
Intrinsic value (pretest)	0.08	(0.04)	†							0.63	(0.02)	***									
Utility value (pretest)	0.04	(0.03)											0.53	(0.03)	***						
Dispositional interest (pretest)	−0.01	(0.04)														0.65	(0.03)	***			
Competence beliefs (pretest)	0.06	(0.04)																	0.61	(0.03)	***
Achievement (pretest)	0.21	(0.02)	***	0.03	(0.03)		−0.03	(0.02)		0.04	(0.02)		0.07	(0.02)	**	0.02	(0.02)		0.05	(0.02)	*
Science grade	0.12	(0.03)	***	0.14	(0.02)	***	−0.21	(0.03)	***	0.10	(0.02)	***	0.14	(0.02)	***	0.08	(0.02)	***	0.15	(0.03)	***
<i>Class level</i>																					
School only	0.50	(0.14)	**	0.00	(0.08)		0.02	(0.07)		0.04	(0.06)		−0.09	(0.07)		0.05	(0.08)		0.00	(0.06)	
SCOL & school	0.50	(0.12)	***	−0.02	(0.07)		−0.19	(0.07)	**	0.04	(0.06)		−0.10	(0.07)		0.04	(0.06)		0.15	(0.06)	**
SCOL only	0.17	(0.14)		0.04	(0.08)		−0.14	(0.06)	*	0.07	(0.07)		−0.13	(0.07)	†	−0.03	(0.07)		0.11	(0.05)	*
<i>R<sup>2</sup></i>																					
Student level	0.13			0.40			0.29			0.45			0.34			0.46			0.49		
Class level	0.32			0.68			0.70			0.80			0.84			0.60			0.91		
Snijders & Bosker	0.25			0.41			0.30			0.45			0.34			0.46			0.50		
N	1452			1464			1467			1467			1462			1468			1464		

Note. At the student level, the covariate (task value at pretest) was group-mean centered. Manifest aggregation was used for the class-level predictor (Marsh et al., 2009). At the class level, the covariate was grand-mean centered. All continuous variables were z-standardized beforehand. The reference group for the treatment was the control group.  $\beta$  = standardized regression coefficient. DV = dependent variable. Snijders & Bosker  $R^2$ : mean squared prediction error following Snijders and Bosker (1994).

†  $p < 0.10$ .  
 \*  $p < 0.05$ .  
 \*\*  $p < 0.01$ .  
 \*\*\*  $p < 0.001$ .

competence beliefs at posttest (Table 5, Fig. 4). In other words, we found some but not very strong support for intervention effects on motivational outcomes. It is important to note that these positive effects were found only in the two intervention conditions that included a SCOL visit.

### 3.5. Treatment effects on achievement

The effects of the lab-work conditions on achievement were also tested via multilevel analyses; the control group served as the reference group. Trait motivation, achievement, and science

**Table 6**

Effects of the lab-work conditions on students' learning achievement, subjective task values, dispositional interest, and competence beliefs (follow-up).

	Achievement			Attainment value			Cost			Intrinsic value			Utility value			Dispositional interest			Competence beliefs		
	$\beta$	S.E.	p	$\beta$	S.E.	p	$\beta$	S.E.	p	$\beta$	S.E.	p	$\beta$	S.E.	p	$\beta$	S.E.	p	$\beta$	S.E.	p
<i>Student level</i>																					
Attainment value (pretest)	0.01	(0.04)		0.55	(0.04)	***															
Cost (pretest)	−0.03	(0.03)					0.40	(0.03)	***												
Intrinsic value (pretest)	0.01	(0.05)								0.62	(0.03)	***									
Utility value (pretest)	0.08	(0.04)	*										0.55	(0.03)	***						
Dispositional interest (pretest)	0.05	(0.04)														0.59	(0.03)	***			
Competence beliefs (pretest)	−0.03	(0.04)																	0.56	(0.03)	***
Achievement (pretest)	0.27	(0.03)	***	0.03	(0.02)		−0.04	(0.02)	†	0.01	(0.02)		0.05	(0.02)	†	0.02	(0.02)		0.04	(0.02)	*
Science grade	0.13	(0.04)	**	0.11	(0.03)	***	−0.17	(0.04)	***	0.08	(0.03)	**	0.09	(0.02)	***	0.11	(0.02)	***	0.12	(0.03)	***
<i>Class level</i>																					
School only	0.65	(0.19)	**	0.07	(0.08)		−0.01	(0.09)		0.02	(0.09)		−0.09	(0.07)		0.00	(0.10)		0.06	(0.09)	
SCOL & school	0.38	(0.16)	*	0.04	(0.08)		−0.17	(0.07)	*	0.10	(0.08)		−0.03	(0.07)		0.08	(0.10)		0.14	(0.08)	†
SCOL only	0.20	(0.18)		0.17	(0.09)	†	−0.04	(0.06)		0.09	(0.10)		−0.01	(0.07)		0.10	(0.10)		0.06	(0.09)	
$R^2$																					
Student level	0.17			0.36			0.24			0.42			0.34			0.41			0.40		
Class level	0.34			0.62			0.83			0.59			0.92			0.75			0.87		
Snijders & Bosker	0.25			0.37			0.25			0.41			0.35			0.43			0.41		
N	1452			1464			1467			1467			1462			1468			1464		

Note. At the student level, the covariate (task value at pretest) was group-mean centered. Manifest aggregation was used for the class-level predictor (Marsh et al., 2009). At the class level, the covariate was grand-mean centered. All continuous variables were z-standardized beforehand. The reference group for the treatment was the control group.  $\beta$  = Standardized regression coefficient. DV = dependent variable. Snijders & Bosker  $R^2$ : mean squared prediction error following Snijders and Bosker (1994).

†  $p < 0.10$ .\*  $p < 0.05$ .\*\*  $p < 0.01$ .\*\*\*  $p < 0.001$ .

grade were included as covariates. One week after the intervention (posttest), there were significant effects of achievement when comparing the *School only* and *SCOL & school* conditions with the control condition (Table 5, Fig. 3). Six weeks after the intervention (follow-up), these positive effects for the *School only* and *SCOL & school* conditions remained but were less pronounced (Table 5). Moreover, there were significant differences between the lab-work conditions: The two conditions involving lessons at school (*School only* and *SCOL & school*) resulted in significantly higher achievement than the *SCOL only* condition, and the difference between *School only* and *SCOL only* remained on the follow-up test (Fig. 3). Even though the *SCOL only* condition had performed significantly better than the control at pretest, this group did not differ from the control condition at posttest<sup>3</sup> and follow-up.

The effect sizes (Snijders & Bosker's  $R^2$ ) were 0.13–0.23 for the motivational state variables (Table 4), 0.10–0.15 for the theoretical part, and 0.12–0.20 for the practical part. Effect sizes were 0.30–0.50 for the motivational trait variables (posttest; Table 5) and 0.25–0.43 at follow-up (Table 5). The achievement effect sizes were the same at posttest and follow-up (0.25). To summarize, the effect sizes for the state components of motivation were higher for the practical part than for the theoretical part; the effect sizes for the trait components of motivation were higher on the posttest than on the follow-up test.

<sup>3</sup> The detailed treatment effects on achievement (posttest), partitioned into five achievement subtests, were reported in Itzek-Greulich et al. (2015). To add to these results, the detailed results for the follow-up are shown in Appendix A. Regarding Chemical Terms, the three treatment conditions were more successful than the control group. The students from the *School only* condition had higher achievement in Carbohydrate Specific Knowledge than the other two conditions and the control. Moreover, the *School only* condition had higher achievement than the control group in Chemical Analysis and Declarative Knowledge. As a side note, the covariates in Itzek-Greulich et al. (2015) differ from those implemented in the current study.

#### 4. Discussion

The present study examined the effectiveness of different lab-work learning environments with regard to motivational and achievement outcomes. Three lab-work learning arrangements (*School only*, *SCOL & school*, *SCOL only*) and a control group were compared via a cluster randomized controlled trial. We assessed achievement outcomes, and relying on the EVT by Eccles and Wigfield (2002), we assessed state and trait components of motivation.

*Discussion of Hypothesis 1, state motivation.* In line with the assumption that lab-work is favorable for student motivation, the findings indicated that the three lab-work conditions were successful in comparison with the control group with regard to the state measures. In fact, students in the three lab-work conditions had more positive state motivational outcomes on all five state measures than their counterparts in the control condition. The pattern of findings is also in line with the assumption (Dohn, 2011; Eccles, 2005; Krapp & Prenzel, 2011) that certain aspects of learning environments have a direct impact on state motivational measures. More specifically, students in the intervention conditions experienced more joy, higher situational interest, higher situational competence beliefs, and lower boredom (*School only* and *SCOL & school*) than the students in the control condition. Interestingly, there were no significant differences in state motivation between the three lab-work conditions, except for situational competence. This finding suggests that “what” students do is most important for their momentary experience of motivation, not “where” they do it: Doing lab-work instead of engaging in regular learning units is more motivating (“hands-on” and “minds-on”; Abrahams & Reiss, 2012), but the in-school versus out-of-school difference did not matter. Concerning situational competence, supposedly the better equipped out-of-school lab increased the students' situational competence, but only when the theoretical part of the lesson took place at the school.



*Discussion of Hypothesis 2, trait motivation.* The pattern of results was less pronounced for trait motivation. We found positive effects of the lab-work conditions for two of the six outcomes, and these positive effects on trait motivation were found only in the conditions involving out-of-school learning (SCOL & school and SCOL only). The results were somewhat different with regard to the more trait-like measures, where we found fewer effects of the intervention conditions. This is in line with [Abrahams \(2009\)](#), according to which practical work is limited to short-term motivational gains. In addition, the positive effects were restricted to the two conditions that included a SCOL visit. In general, it is well-known that positive momentary motivational states do not automatically translate into more stable effects; for instance, the person-object theory of motivation ([Krapp et al., 1992](#)) distinguishes between four stages in the accumulation of interest as a stable personal feature ([Hidi & Renninger, 2006](#)). For this reason, the pattern of findings might not be all that surprising. However, it is interesting to see that cost and competence beliefs were positively affected by the SCOL conditions. Therefore, the change in the learning environment (SCOL visit) was associated with some change in trait motivation.

*Discussion of Hypothesis 3, achievement.* Regarding achievement and comparing the effects on a posttest and a follow-up test, the posttest effects were still present 6 weeks after treatment in the *School only* and *SCOL & school* conditions. However, 6 weeks after the intervention, the positive effects on achievement and motivation were fewer and smaller than directly after the intervention. This goes in line with other studies that tested for treatment effects on post- and follow-up-tests (e.g., [Price, Lee, Subbarao, Kasal, & Aguilera, 2015](#)). Regarding inquiry-based learning, the 1-day course at the SCOL may be too short for formulating and implementing a complete science project that promotes the principles of scientific research. Especially for inquiry-based learning in a science class, continuity and sustainable development of learning is given only when the SCOL visit is thoroughly embedded in a triad of preparation, experimentation, and debriefing ([Kisiel, 2014](#); [Tal, Bamberger, & Morag, 2005](#)), for example, to reduce the “novelty space” ([Orion & Hofstein, 1994](#)).

#### 4.1. Diverging effects of the lab-work settings on state motivation, trait motivation, and achievement

The pattern of effects is quite interesting as it shows some differential findings with regard to the three broad targets of the intervention: The intervention had a positive effect on situational interest (state), whereas there was a less pronounced treatment effect on dispositional interest (trait). This is in line with [Knogler, Harackiewicz, Gegenfurtner, and Lewalter's \(2015\)](#) finding that situational interest is sensitive to situational circumstances. The regular school setting was more effective at enhancing achievement, the SCOL was more effective at enhancing trait motivation, and all kinds of lab-work were associated with an increase in state motivation no matter whether it occurred in or out of school – except for situational competence. Accordingly, for the achievement outcomes, it seems important to include the highly structured school element, whereas, for motivational change, a SCOL visit might be particularly important.

The smaller variation (ICC) between classes for the motivational trait variables indicates that these variables measure more stable constructs that are not affected on the class level (from the treatment or the teacher) as much as the state variables and achievement. The higher amplitude in variation between classes for the state outcomes and for achievement suggests differences between the treatment groups or between classes in the same treatment group, and this indicates qualitative differences in the

implementation of the treatments. In fact, the differences were highest in the classes in the *SCOL & school* and *SCOL only* groups. This is interesting because, at the SCOL, classes were instructed by only two different instructors, whereas the classes in the *School only* group were taught by their respective science teachers.

#### 4.2. Practical implications

Experimental investigation is part of scientific literacy and is an important educational aim ([NRC, 2012](#)). Here, it is especially interesting to look at adolescents at the age of 15 years because at this age, students realize whether they are interested in studying science ([Venville, Rennie, Hanbury, & Longnecker, 2013](#)). Therefore, students should have the opportunity to get involved in scientific experimentation, and a SCOL offers authentic hands-on activities in a scientific setting that, as seen in the results of our study, holds additional longer-lasting motivational value for students. Therefore, experimental project work at a SCOL can be seen as a worthwhile supplement to a good STEM education. Overall, the results clearly indicate that teachers should be asked to think about embedding a SCOL visit into the school curriculum. However, teachers have to be aware that the SCOL visit per se does not enhance achievement when it is isolated and not carefully prepared and students are not debriefed on it at school. In other words, experimental investigation should not be undertaken at the expense of effectively conveying the underlying concepts and purposes. Thus, students' science achievement is likely to be improved by adding a wider range of experimental and methodological designs ([Hausmann, 2012](#); [Pawek, 2012](#)) in a SCOL or school setting. This implies that motivation can increase if teachers use a variety of teaching methods (e.g., psychomotor activity in student-driven experiments). However, carefully thought out organizational planning is required when visiting a SCOL to increase the expected gains in motivation and achievement. Moreover, the many new experiences and challenges of a SCOL visit may strain students cognitively, and this cognitive overload should be debriefed back at school.

An aspect that we were not able to experimentally manipulate in the present study is the professional provenance of the scientists at the SCOL, which may have had an impact on their teaching style. There is no certification process yet, and *SCOL scientist* is not a separate professional group. SCOL scientists come from different careers: Some are science teachers, but most of them are scientists with no teaching experience. Therefore, further training for SCOL scientists might be needed to strengthen the scientists' pedagogical competence and to ensure that students will have meaningful learning experiences with them. When the SCOL scientists have pedagogical training, a SCOL visit might also produce higher gains in achievement. Also, a closer and longer-lasting cooperation between teachers and SCOL scientists would be fruitful for the integration of a SCOL into the school curriculum. In fact, several teachers might not be adequately prepared when they take their class to an out-of-school learning environment ([Griffin & Symington, 1997](#)) because students and teachers may have the wrong expectations of a SCOL visit ([Garner & Eilks, 2015](#)). We believe that research such as the study presented here has the potential to contribute to increasing the effects of SCOL visits.

#### 4.3. Limitations

The present study contributes to research on emotions during lab-work and the motivational outcomes. The study compared three treatment groups, including a combined setting: the *SCOL*

& school group. We distinguished between state and trait components of motivation by using a large sample of school classes that were randomly assigned to the treatment conditions from a fairly large sample of ninth-grade students. At the same time, however, some limitations have to be addressed when interpreting the results.

First and most relevant, the degree to which our findings are comparable across different topics and different SCOLs is unclear. We used a control group that can best be characterized as a wait-list control group. Such a choice is very common when comparing the effectiveness of an intervention in real-world contexts. At the same time, the differences with respect to the control condition might not have been due to the lab-work but may also have depended on the topic of the intervention. The topic was organic chemistry, which may have elicited more favorable outcomes than an inorganic topic, and thus we believe that future research is needed to compare practical lab-work on starch chemistry with other means of teaching and learning (e.g., teacher-centered or theoretical group work) and should also include control groups in which the topic of starch is taught “as usual.” Unfortunately, our study is one of the first large group-randomized trials to consider an integrative SCOL & school condition and to use several predictors of achievement, state motivation, and trait motivation. Given the dearth of research in this area and information on its practical relevance, more research is needed.

Second, in a related vein, standardizing the contents and how the SCOL was incorporated into the curriculum facilitated an unbiased comparison of the treatment conditions. Another approach would have been to apply online tools or to allow the teachers to decide how to handle the preparation and debriefing at school. Again, more studies are necessary to investigate the relative benefits.

Third, the visit was brief (eight lessons) and was not conducted on a regular basis. Many curricular school units devote more time than we did, and it would be interesting to see whether the positive boost in motivation that occurred during and after the SCOL visit could be ascribed solely to the novelty effect or whether the positive motivational effects could be found if SCOL visits were implemented on a regular basis. However, SCOL lab units of more than 8 h are quite atypical, and thus, the setting used in the present study has high ecological validity. Moreover, these 8 h were sufficient for providing differentiated and noticeably positive motivational feedback in all treatment conditions when compared with the control group, yet with smaller gains in trait motivation than in state motivation, supposedly because the intervention was too short.

Fourth, SCOLs might be better at conveying practical/psychomotor skills and aspects that are not easily measured with paper-and-pencil achievement tests. For instance, research could use qualitative methods such as video analyses, for example, to investigate the collaborative and communicative processes of practical scientific work (see, e.g., Högström, Ottander, & Benckert, 2010; Schmidt, DiFuccia, & Ralle, 2014). However, the quantitative approach of our study was a premise for the multilevel analyses of the hierarchically structured data, and with a large sample size, the application of standardized tests was the most convenient approach for testing students' achievement.

Fifth, the teacher is familiar with the students and has the advantage of knowing how to present and teach contents didactically to make sure their students pick up the contents quickly, whereas scientists at SCOLs lack this knowledge, resulting in a “bias” toward the teachers (Howitt, Rennie, Heard, & Yunken, 2009). However, in the study presented here, the conception of

treatments, specifically the learning arrangements, are real teaching methods and therefore have high practicability.

Sixth, in our study, the typically better equipment at the SCOL did not play a role because the same experimental kits were given to the participating classes in the *School only* group to balance this usual SCOL advantage. The chosen topic had to be relatively easy so that all treatment groups could implement the intervention (e.g., regarding the safety precautions and the equipment). A highly sophisticated experiment such as DNA inheritance and genes (Dairianathan & Subramaniam, 2011) or nanotechnology (Schwarzer, Akaygun, Sagun-Gokoz, Anderson, & Blonder, 2015) could not have been implemented with our study design.

#### 4.4. Conclusion

The present study applied a wide range of motivational measures (state emotions, state motivation, and trait motivation) to test whether lab-work at school or in an out-of-school learning arrangement (here, at a SCOL) are equally effective for raising achievement and motivational outcomes. We found positive effects of lab-work education on state motivation, achievement (in-school learning), and—although less pronounced—trait measures of motivation (out-of-school learning). Thus, the student-centered approach of experimentation in an out-of-school setting is a feasible way to complement teaching at school when the out-of-school learning is thoroughly prepared and the students are debriefed at school. Focusing on out-of-school science learning at science centers (SCOLs), a topic that has been largely overlooked in the last few decades by researchers, our study highlights differential effects of the same intervention in terms of different outcomes: The lessons at school were more effective for producing gains in achievement, whereas the lessons at the out-of-school lab were more effective at establishing longer-lasting motivation, which could be attributed to the impact of the relatively new learning facility of the science center with its professional equipment. Since research has only recently begun to investigate the settings and predictors of out-of-school learning, future research will need to explore emotional, motivational and achievement-related aspects in more detail and may wish to focus on how to enhance students interests and abilities.

#### Author note

Heike Itzek-Greulich was a member of the “Cooperative Research Training Group” of the University of Education, Ludwigsburg, and the University of Tübingen; the group was funded by the Ministry of Science, Research, and the Arts in Baden-Württemberg. This research was also funded by the German Funds of the Chemical Industry (Fonds der chemischen Industrie) Frankfurt, the EXPERIMENTA gGmbH Heilbronn, and the University of Education Ludwigsburg (Forschungsförderung).

#### Acknowledgments

This research was funded by the Ministry of Science, Research, and the Arts in Baden-Württemberg. For financial support, we would also like to thank The German Funds of the Chemical Industry (Fonds der chemischen Industrie, Förderlinie Chemiedidaktik) regarding materials for the school lessons (treatment groups SCOL & school and School only), the EXPERIMENTA gGmbH and University of Education Ludwigsburg (Forschungsförderung).

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

**Appendix A. Detailed effects of the intervention on achievement outcomes (follow-up)**

	Carbohydrate Specific Knowledge			Chemical Analysis			Chemical Terms			Experimental Specific Knowledge			Declarative Knowledge		
	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>	$\beta$	S.E.	<i>p</i>
<i>Student level</i>															
Gender (1 = female)	0.15	(0.07)	*	0.14	(0.07)	*	−0.09	(0.07)		0.16	(0.05)	**	0.21	(0.06)	***
Pretest scores (DV)				0.13	(0.04)	***	0.38	(0.03)	***	0.27	(0.03)	***	0.29	(0.03)	***
Language spoken at home	0.00	(0.10)		0.10	(0.06)		0.19	(0.07)	**	0.25	(0.07)	***	0.18	(0.06)	**
<i>Class level</i>															
<i>SCOL &amp; school as reference group</i>															
School only	0.39	(0.11)	***	0.20	(0.11)	†	0.03	(0.08)		0.07	(0.15)		0.26	(0.14)	†
SCOL only	−0.09	(0.13)		0.03	(0.07)		−0.04	(0.08)		0.07	(0.16)		−0.01	(0.14)	
Control condition	−0.06	(0.14)		−0.05	(0.08)		−0.33	(0.08)	***	−0.25	(0.18)		−0.29	(0.15)	*
<i>Control as reference group</i>															
School only	0.45	(0.15)	**	0.25	(0.12)	*	0.36	(0.09)	***	0.32	(0.20)		0.55	(0.15)	***
SCOL & school	0.06	(0.14)		0.05	(0.08)		0.33	(0.08)	***	0.25	(0.18)		0.29	(0.15)	*
SCOL only	−0.04	(0.16)		0.09	(0.09)		0.29	(0.09)	**	0.32	(0.21)		0.29	(0.15)	†
<i>School only as reference group</i>															
SCOL & school	−0.39	(0.11)	***	−0.20	(0.11)	†	−0.03	(0.08)		−0.07	(0.15)		−0.26	(0.14)	†
SCOL only	−0.49	(0.13)	***	−0.17	(0.12)		−0.07	(0.10)		0.00	(0.19)		−0.27	(0.14)	†
Control condition	−0.45	(0.15)	**	−0.25	(0.12)	*	−0.36	(0.09)	***	−0.32	(0.20)		−0.55	(0.15)	***
<i>R<sup>2</sup> (SCOL &amp; school as reference group)</i>															
Student level	0.01			0.02			0.15			0.10			0.11		
Class level	0.34			0.52			0.77			0.17			0.32		
Snijders & Bosker	0.04			0.03			0.18			0.12			0.16		
ICC	0.11			0.03			0.05			0.19			0.17		
<i>N</i>	1601			1594			1617			1617			1634		

Note. DV, dependent variable at pretest; gender was coded: 1 = female, 0 = male; language spoken at home was coded: 1 = German, 0 = non German; Snijders & Bosker R<sup>2</sup>: mean squared prediction error following Snijders and Bosker (1994); Carbohydrate Specific Knowledge was not measured prior to the intervention (pretest). ICC = intraclass correlation.

\*  $p < 0.05$ .

\*\*  $p < 0.01$ .

\*\*\*  $p < 0.001$ .

**Appendix B. Goodness of fit statistics for alternative models of longitudinal confirmatory analyses**

Dimension	Hypothesis	CFI	RMSEA	$\Delta$ CFI	$\Delta$ RMSEA
Attainment (4 items)	(1) Configural invariance	0.922	0.067		
	(2) Metric invariance	0.919	0.065	(1–2) 0.003	(1–2) 0.002
	(3) Scalar invariance	0.903	0.067	(2–3) 0.016	(2–3) −0.002
Cost (2 items)	(1) Configural invariance	0.995	0.030		
	(2) Metric invariance	0.992	0.034	(1–2) 0.003	(1–2) −0.004
	(3) Scalar invariance	0.979	0.045	(2–3) 0.013	(2–3) −0.011
Intrinsic (5 items)	(1) Configural invariance	0.914	0.063		
	(2) Metric invariance	0.910	0.062	(1–2) 0.004	(1–2) 0.001
	(3) Scalar invariance	0.899	0.063	(2–3) 0.011	(2–3) −0.001
Utility (3 items)	(1) Configural invariance	0.926	0.073		
	(2) Metric invariance	0.917	0.071	(1–2) 0.009	(1–2) 0.002
	(3) Scalar invariance	0.909	0.068	(2–3) 0.008	(2–3) 0.003
Dispositional interest (5 items)	(1) Configural invariance	0.849	0.091		
	(2) Metric invariance	0.842	0.089	(1–2) 0.007	(1–2) 0.002
	(3) Scalar invariance	0.827	0.089	(2–3) 0.015	(2–3) <0.001

## Appendix B (continued)

Dimension	Hypothesis	CFI	RMSEA	$\Delta$ CFI	$\Delta$ RMSEA
Competence beliefs (6 items)	(1) Configural invariance	0.909	0.059		
	(2) Metric invariance	0.903	0.059	(1–2) 0.006	(1–2) <0.001
	(3) Scalar invariance	0.897	0.059	(2–3) 0.006	(2–3) <0.001
Achievement (33 items)	(1) Configural invariance	0.209	0.053		
	(2) Metric invariance	0.205	0.053	(1–2) 0.004	(1–2) <0.001
	(3) Scalar invariance	0.171	0.054	(2–3) 0.034	(2–3) –0.001

Note. When the decrease in fit was less than 0.01 for CFI and when the RMSEA differed by less than 0.015, the more restrictive model was preferred.

## References

- Abrahams, I. (2009). Does practical work really motivate? A study of the affective value of practical work in secondary school science. *International Journal of Science Education*, 31(17), 2335–2353. <http://dx.doi.org/10.1080/09500690802342836>.
- Abrahams, I., & Reiss, M. J. (2012). Practical work: Its effectiveness in primary and secondary schools in England. *Journal of Research in Science Teaching*, 49, 1035–1055.
- Adams, J. D., Gupta, P., & DeFelice, A. (2012). Schools and informal science settings: Collaborate, co-exist, or assimilate? *Cultural Studies of Science Education*, 7, 409–416.
- Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology*, 36(1), 4–12. <http://dx.doi.org/10.1016/j.cedpsych.2010.08.001>.
- Arbuckle, L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum.
- Ateşkan, A., & Lane, J. F. (2016). Promoting field trip confidence: Teachers providing insights for pre-service education. *European Journal of Teacher Education*, 1–12. <http://dx.doi.org/10.1080/02619768.2015.1113252>.
- Behrendt, M., & Franklin, T. (2014). A review of research on school field trips and their value in education. *International Journal of Environmental and Science Education*, 9(3), 235–245.
- Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, 94, 577–616.
- Bolstad, R., Bull, A., Carson, S., Gilbert, J., & MacIntyre, B. (2013). *Strengthening engagements between schools and the science community: Final report*. Wellington: New Zealand Council for Educational Research.
- Brandt, A., Möller, J., & Kohse-Höinghaus, K. (2008). Was bewirken außerschulische Experimentierlabors? [What's the effect of science laboratories?]. *Zeitschrift für Pädagogische Psychologie*, 22, 5–12.
- Braund, M., & Reiss, M. (2006). Towards a more authentic science curriculum: The contribution of out-of-school learning. *International Journal of Science Education*, 28, 1373–1388.
- Butz, W. P. (2004). *Will the scientific and technical workforce meet the requirements of the federal government?* Santa Monica, CA: RAND.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48, 1–29.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 233–255.
- Chow, A., Eccles, J. S., & Salmela-Aro, K. (2012). Task value profiles across subjects and aspirations to physical and IT-related sciences in the United States and Finland. *Developmental Psychology*, 48(6), 1612–1628.
- Committee on Prospering in the Global Economy of the 21st Century (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington: National Academies Press.
- Dairianathan, A., & Subramaniam, R. (2011). Learning about inheritance in an out-of-school setting. *International Journal of Science Education*, 33, 1079–1108.
- DeWitt, J., & Storksdiel, M. (2008). A short review of school field trips: Key findings from the past and implications for the future. *Visitor Studies*, 11, 181–197.
- Dohn, N. B. (2011). Situational interest of high school students who visit an aquarium. *Science Education*, 95, 337–357.
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 74–146). San Francisco, CA: W. H. Freeman.
- Eccles, J. S. (2005). Subjective task values and the Eccles et al. model of achievement related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105–121). Guilford Press.
- Eccles, J. S. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco: Freeman.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology* (pp. 1017–1095). Hoboken, NJ, US: John Wiley & Sons Inc.
- Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences*. New York: Guilford Press.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138.
- European Commission (2010). *Special eurobarometer 73.1. January 2010–February 2010. Science and technology. Report*.
- Fallik, O., Rosenfeld, S., & Eylon, B.-S. (2013). School and out-of-school science: A model for bridging the gap. *Studies in Science Education*, 49, 69–91.
- Garner, N., & Eilks, I. (2015). The expectations of teachers and students who visit a nonformal student chemistry laboratory. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 1197–1210. <http://dx.doi.org/10.12973/eurasia.2015.1414a>.
- Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B. M., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, 663–677. <http://dx.doi.org/10.1037/edu0000003>.
- Gibson, H. L., & Chase, C. (2002). Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science. *Science Education*, 86, 693–705.
- Glowinski, I., & Bayrhuber, H. (2011). Student labs on a university campus as a type of out-of-school learning environment: Assessing the potential to promote students' interest in science. *International Journal of Environmental and Science Education*, 6, 371–392.
- Gomes, D., & McCauley, V. (2012). Science outreach and science education at primary level in Ireland: A mixed methods study. *New perspectives in science education*.
- Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93, 3–13.
- Griffin, J., & Symington, D. (1997). Moving from task-oriented to learning-oriented strategies on school excursions to museums. *Science Education*, 81, 763–779.
- Guderian, P., Priemer, B., & Schön, L.-H. (2006). In den Unterricht eingebundene Schülerlaborbesuche und deren Einfluss auf das aktuelle Interesse an Physik [Embedded out-of-school lab visits and their impact on situational interest in science subjects]. *PhyDid A – Physik und Didaktik in Schule und Hochschule*, 2, 142–149.
- Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments? *Journal of Research in Science Teaching*, 37, 655–675.
- Hausmann, D. (2012). Extracurricular science labs for STEM talent support. *Roeper Review*, 34, 170–182.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41, 111–127.
- Hoffmann, L. (2002). Promoting girls' interest and achievement in physics classes for beginners. *Learning and Instruction*, 12, 447–465.
- Hofstein, A., Eilks, I., & Bybee, R. W. (2010). Societal issues and their importance for contemporary science education. In I. Eilks & B. Ralle (Eds.), *Contemporary science education* (pp. 5–22). Aachen: Shaker.
- Hofstein, A., Kesner, M., & Ben-Zvi, R. (1999). Student perceptions of industrial chemistry classroom learning environments. *Learning Environments Research*, 2, 291–306.
- Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42, 791–806.
- Högström, P., Ottander, C., & Benckert, S. (2010). Lab work and learning in secondary school chemistry: The importance of teacher and student interaction. *Research in Science Education*, 40(4), 505–523. <http://dx.doi.org/10.1007/s11165-009-9131-3>.



- Howitt, C., Rennie, L., Heard, M., & Yuncken, L. (2009). The scientists in schools project. *Teaching Science*, 55, 35–38.
- Itzek-Greulich, H., Flunger, B., Vollmer, C., Nagengast, B., Rehm, M., & Trautwein, U. (2015). Effects of a science center outreach lab on school students' achievement – Are student lab visits needed when they teach what students can learn at school? *Learning and Instruction*, 38, 43–52. <http://dx.doi.org/10.1016/j.learninstruc.2015.03.003>.
- Itzek-Greulich, H., Flunger, B., Vollmer, C., Nagengast, B., Rehm, M., & Trautwein, U. (2014). The impact of a science center outreach lab workshop on German 9th graders' achievement in science. In ESERA (Ed.), *10th Conference of the European Science Education Research Association, proceedings* (pp. 97–106).
- Jarvis, T., & Pell, A. (2005). Factors influencing elementary school children's attitudes toward science before, during, and after a visit to the UK National Space Centre. *Journal of Research in Science Teaching*, 42, 53–83.
- Jensen, E. A., & Lister, T. J. (2016). Evaluating indicator-based methods of 'measuring long-term impacts of a science center on its community' (Comment). *Journal of Research in Science Teaching*, 53(1), 60–64. <http://dx.doi.org/10.1002/tea.21297>.
- King, D., & Ritchie, S. M. (2012). Learning science through real-world contexts. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education: Pt. 1. Second international handbook of science education* (pp. 69–79). Dordrecht: Springer.
- Kisiel, J. F. (2014). Clarifying the complexities of school-museum interactions: Perspectives from two communities. *Journal of Research in Science Teaching*, 51(3), 342–367. <http://dx.doi.org/10.1002/tea.21129>.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., & Prenzel, M. (Eds.). (2010). *PISA 2009. Bilanz nach einem Jahrzehnt [PISA 2009. Results after a decade]*. Münster: Waxmann.
- Knogler, M., Harackiewicz, J. M., Gegenfurtner, A., & Lewalter, D. (2015). How situational is situational interest?: Investigating the longitudinal structure of situational interest. *Contemporary Educational Psychology*, 43, 39–50. <http://dx.doi.org/10.1016/j.cedpsych.2015.08.004>.
- Krapp, A. (2000). Interest and human development during adolescence: An educational-psychological approach. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 109–128). London: Elsevier.
- Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning and development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3–25). Hillsdale, NJ: Erlbaum.
- Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, 33, 27–50.
- Lau, S., & Roeser, R. W. (2002). Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science. *Educational Assessment*, 8, 139–162.
- Luehmann, A. L. (2009). Students' perspectives of a science enrichment programme: Out-of-school inquiry as access. *International Journal of Science Education*, 31, 1831–1855.
- Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) III: A theoretical and empirical basis for the measurement of multiple dimensions of late adolescent self-concept: An interim test manual and a research monograph*. Macarthur, New South Wales, Australia: University of Western Sydney, Faculty of Education.
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. O., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44, 764–802.
- McClafferty, T., & Rennie, L. (1993). Learning in science centres and science museums: A review of recent studies. *Research in Science Education*, 23, 351.
- Meissner, B., & Bogner, F. X. (2011). Enriching students' education using interactive workstations at a salt mine turned science center. *Journal of Chemical Education*, 88, 510–515.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543.
- Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology*, 85, 424–436.
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nielsen, W. S., Nashon, S., & Anderson, D. (2009). Metacognitive engagement during field-trip experiences: A case study of students in an amusement park physics program. *Journal of Research in Science Teaching*, 46(3), 265–288. <http://dx.doi.org/10.1002/tea.20266>.
- NRC (Ed.). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
- Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of Research in Science Teaching*, 31, 1097–1119.
- Osborne, J., & Dillon, J. (2008). *Science education in Europe: Critical reflections. A report to the Nuffield Foundation*. London: King's College.
- Patrick, P. G., Mathews, C., & Tunnicliffe, S. D. (2013). Using a field trip inventory to determine if listening to elementary school students' conversations, while on a zoo field trip, enhances preservice teachers' abilities to plan zoo field trips. *International Journal of Science Education*, 35, 2645–2669.
- Pawek, C. (2012). Schülerlabore als interessensfördernde außerschulische Lernumgebungen [Science laboratories as a supporting interest environment]. In D. Brovelli, K. Fuchs, R. v. Niederhäusern, & A. v. Rempfler (Eds.), *Kompetenzentwicklung an Außerschulischen Lernorten* (pp. 69–94). Münster/Wien/Zürich: LIT.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18, 315–341.
- Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, 98, 583–597.
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36, 36–48.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37, 91–105.
- Pekrun, R., Götz, T., Zirnigib, A., & Jullien, S. (2002). *PALMA: Projekt zur Analyse der Leistungsentwicklung in Mathematik [PALMA – Project for the analysis of achievement development in mathematics]*. Universität München, Institut Pädagogische Psychologie.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, 25(1), 92–104. <http://dx.doi.org/10.1006/ceps.1999.1017>.
- Pluth, M. D., Boettcher, S. W., Nazin, G. V., Greenaway, A. L., & Hartle, M. D. (2015). Collaboration and near-peer mentoring as a platform for sustainable science education outreach. *Journal of Chemical Education*, 92(4), 625–630. <http://dx.doi.org/10.1021/ed500377m>.
- Price, C. A., Lee, H., Subbarao, M., Kasal, E., & Aguilera, J. (2015). Comparing short- and long-term learning effects between stereoscopic and two-dimensional film at a planetarium. *Science Education*, 99. <http://dx.doi.org/10.1002/sce.21185>.
- Priemer, B., & Pawek, C. (2014). Out-of-school STEM learning in Germany: Can we catch and hold students' interest? *Paper presented at the NARST 2014 annual conference*.
- Pugh, K. J. (2011). Transformative experience: An integrative construct in the spirit of Deweyan pragmatism. *Educational Psychologist*, 46(2), 107–121. <http://dx.doi.org/10.1080/00461520.2011.558817>.
- Randler, C., Ilg, A., & Kern, J. (2005). Cognitive and emotional evaluation of an amphibian conservation program for elementary school students. *Journal of Environmental Education*, 37, 43–52.
- Rennie, L. (2014). Learning science outside of school. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. II, pp. 120–144). London: Routledge.
- Rennie, L., & McClafferty, T. (1995). Using visits to interactive science and technology centres, museums, aquaria, and zoos to promote learning in science. *Journal of Science Teacher Education*, 6(4), 175–185.
- Rickinson, M., Dillon, J., Teamey, K., Morris, M., Choi, M. Y., Sanders, D., & Benefield, P. (2004). *A review of research on outdoor learning*. Shrewsbury: Field Studies Council.
- Scharfenberg, F.-J., Bogner, F. X., & Klautke, S. (2007). Learning in a gene technology laboratory with educational focus: Results of a teaching unit with authentic experiments. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 35, 28–39.
- Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 197–222). New York/London: Routledge.
- Schmidt, I., DiFuccio, D. S., & Ralle, B. (2014). Science education in out-of-school contexts. In ESERA (Ed.), *10th Conference of the European Science Education Research Association, proceedings* (pp. 3–13).
- Schreiner, C., & Sjøberg, S. (2004). *Sowing the seeds of ROSE: Background, rationale, questionnaire development and data collection for ROSE (The Relevance of Science Education): A comparative study of students' views of science and science education*. Acta didactica. Oslo: Faculty of Education, Department of Teacher Education and School Development (Unipub.).
- Schwan, S., Grajal, A., & Lewalter, D. (2014). Understanding and engagement in places of science experience: Science museums, science centers, zoos, and aquariums. *Educational Psychologist*, 49(2), 70–85. <http://dx.doi.org/10.1080/00461520.2014.917588>.
- Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener [Development an instrument of measuring adolescent self-concept]. *Diagnostica*, 51, 183–194.
- Schwarzer, S., Akaygun, S., Sagun-Gokoz, B., Anderson, S., & Blonder, R. (2015). Using atomic force microscopy in out-of-school settings—Two case studies investigating the knowledge and understanding of high school students. *Journal of Nano Education*, 7(1), 10–27. <http://dx.doi.org/10.1166/jne.2015.1079>.
- Schwichow, M., Zimmerman, C., Croker, S., & Härtig, H. (2016). What students learn from hands-on activities. *Journal of Research in Science Teaching*, 53. <http://dx.doi.org/10.1002/tea.21320> (in press).
- Seybold, B., Braunbeck, T., & Randler, C. (2014). Primate conservation – An evaluation of two different educational programs in Germany. *International Journal of Science and Mathematics Education*, 12, 285–305.
- Singh, M. (2015). A week long summer program does make a difference: A strategy of increasing underrepresented minority students' interest in science. *European Journal of Health and Biology Education*, 4(2), 21–30.
- Sjøberg, S., & Schreiner, C. (2006). How do learners in different cultures relate to science and technology? Results an perspectives from the project ROSE (the Relevance of Science Education). *APFSLT: Asia Pacific Forum of Science Learning and Teaching*, 7, 1–17.

- Snijders, Tom A.B., & Bosker, Roel J. (1994). Modeled variance in two-level models. In: *Sociological Methods & Research* (Vol. 22, pp. S342–S363).
- Stipek, D. J. (1996). Motivation and instruction. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 85–113). New York: Simon and Schuster Macmillan.
- Stuckey, M., Hofstein, A., Mamlok-Naaman, R., & Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education*, 49(1), 1–34. <http://dx.doi.org/10.1080/03057267.2013.802463>.
- Sturm, H., & Bogner, F. X. (2010). Learning at workstations in two different environments: A museum and a classroom. *Studies in Educational Evaluation*, 36, 14–19.
- Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, 49, 515–537.
- Tal, R., Bamberger, Y., & Morag, O. (2005). Guided school visits to natural history museums in Israel: Teachers' roles. *Science Education*, 89, 920–935.
- Tenenbaum, H. R., To, C., Wormald, D., & Pegram, E. (2015). Changes and stability in reasoning after a field trip to a natural history museum. *Science Education*, 99. <http://dx.doi.org/10.1002/sce.21184>.
- Thomas, C. L. (2012). Assessing high school student learning on science outreach lab activities. *Journal of Chemical Education*, 89, 1259–1263.
- Todt, E., & Schreiber, S. (1998). Development of interest. In L. Hoffmann, J. Baumert, A. Krapp, & K. A. Renninger (Eds.), *IPN. Interest and learning. Proceedings of the Seon conference on interest and gender* (Vol. 164, pp. 25–40). Kiel, Germany: IPN.
- Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy-value theory: A latent interaction modeling study. *Journal of Educational Psychology*, 104, 763–777.
- Tytler, R. (2007). *Re-imagining science education: Engaging students in science for Australia's future. Australian education review*. Camberwell, Vic: ACER Press.
- Tytler, R., Osborne, J., Williams, G., Tytler, K., & Cripps Clark, J. (2008). *Opening up pathways: Engagement in STEM across the primary-secondary school transition*. Canberra, ACT: Australian Department of Education, Employment and Workplace Relations.
- Vedder-Weiss, D., & Fortus, D. (2013). School, teacher, peers, and parents' goals emphases and adolescents' motivation to learn science in and out of school. *Journal of Research in Science Teaching*, 50(8), 952–988. <http://dx.doi.org/10.1002/tea.21103>.
- Venville, G., Rennie, L., Hanbury, C., & Longnecker, N. (2013). Scientists reflect on why they chose to study science. *Research in Science Education*, 43(6), 2207–2233. <http://dx.doi.org/10.1007/s11165-013-9352-3>.
- Vierhaus, M., Lohaus, A., & Wild, E. (2016). The development of achievement emotions and coping/emotion regulation from primary to secondary school. *Learning and Instruction*, 42, 12–21. <http://dx.doi.org/10.1016/j.learninstruc.2015.11.002>.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81.
- Wigfield, A., Tonks, S., & Klauda, S. T. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55–75). New York, London: Routledge.
- Willems, A. S. (2011). *Bedingungen des situationalen Interesses im Mathematikunterricht: Eine mehrbenenanalytische Perspektive [Prerequisite factors of the situational interest in mathematics instruction – A multilevel perspective]*. Empirische Erziehungswissenschaft (Vol. 30) Münster, New York, NY, München, Berlin: Waxmann.
- Wünschmann, S., Wüst-Ackermann, P., Randler, C., Vollmer, C., & Itzek-Greulich, H. (2016). Learning achievement and motivation in an out-of-school setting – Visiting amphibians and reptiles in a zoo is more effective than a lesson at school. *Research in Science Education*, 46, 1–22.
- Zhu, X., & Chen, A. (2010). Adolescent expectancy-value motivation and learning: A disconnected case in physical education. *Learning and Individual Differences*, 20, 512–516.