



Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study

Elwin R. Savelsbergh^{a,*}, Gjalt T. Prins^a, Charlotte Rietbergen^b,
Sabine Fechner^{a,1}, Bram E. Vaessen^{a,2}, Jael M. Draijer^a, Arthur Bakker^a

^a Freudenthal Institute, Utrecht University, The Netherlands

^b Faculty of Social and Behavioral Science, Utrecht University, The Netherlands

ARTICLE INFO

Article history:

Received 29 December 2015

Received in revised form 18 July 2016

Accepted 21 July 2016

Available online 26 July 2016

Keywords:

Attitude

Science education

Mathematics education

Motivation

Interest

Meta-analysis

Teaching approach

ABSTRACT

Many teaching approaches have been tried to improve student attitudes and achievement in science and mathematics education. Achievement effects have been synthesized, but a systematic overview of attitude effects is missing. This study provides a meta analytic review based on 56 publications (1988–2014), reporting 65 independent experiments that investigated the effects of teaching approaches on student attitudes in primary or secondary science or mathematics education. Five types of teaching approaches were distinguished: inquiry-based, context-based, computer-based, collaborative learning strategies, and extra-curricular activities. Since many different attitude outcomes were distinguished and attitudes were assessed at different levels of granularity, we did separate analyses for specific and more global outcomes. Outcomes were not significantly different for different educational approaches. When taking all interventions together, significant effects were found for General Attitude ($n = 60$; $d = 0.35$), General Interest ($n = 20$; $d = 0.22$), and Career Interest in Science ($n = 4$; $d = 0.40$). The effects were significantly weaker for studies with older students. Analysis of achievement outcomes yielded a significant and large overall effect ($n = 40$; $d = 0.78$), again with no significant differences between teaching approaches. Although the positive effects might be partly due to novelty, the current findings do counter skepticism about the learning outcomes of interest-oriented teaching approaches.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

In their primary grades, many children demonstrate vivid interest in science and mathematics, but their attitudes decline in the middle grades (Frenzel, Goetz, Pekrun, & Watt, 2010; George, 2006; Kind, Jones, & Barmby, 2007; Osborne, Simon, & Collins, 2003; Potvin & Hasni, 2014a). Interest levels vary by topic and by gender, but the general pattern is quite similar for mathematics and most of the sciences, and the decline occurs in most countries that have attained a certain level of wealth (Osborne et al., 2003; Sjøberg & Schreiner, 2010). The lack of interest in science and mathematics has been on the educational and political agenda for a long time for various reasons, such as the need for a scientifically literate public; the need for employees with a STEM (Science, Technology, Engineering and Mathematics) background; and the need of science itself to

* Corresponding author.

E-mail address: e.r.savelsbergh@uu.nl (E.R. Savelsbergh).

¹ Sabine Fechner is now at University of Paderborn, Germany.

² Bram Vaessen is now at Eindhoven University of Technology, The Netherlands.

gain public support (e.g., AAAS, 1993; European Commission, 2004; Lacey & Wright, 2009; OECD, 2003, 2008; Rocard, Csermley, Jorde, Walberg-Henriksson, & Hemmo, 2007; WRR, 2013).

Many educational innovations, including context-based teaching, inquiry-based teaching, and ICT usage, have been proposed, both in science and mathematics education, to foster positive attitudes, but there is little systematic evidence about which educational approaches are effective to promote interest, attitude, and motivation (Fortus, 2014; Osborne et al., 2003). Moreover, critics have been concerned that such interest-oriented approaches will go at the cost of achievement outcomes (e.g., Kaminski, Sloutsky, & Heckler, 2008; Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004). The purpose of the current study is to gain insight into the magnitude and the robustness of these effects through meta-analytic synthesis of experimental and quasi-experimental studies that investigate the effects of innovative teaching approaches on students' attitudes toward and achievement in science or mathematics.

1. Background

On the basis of theoretical and experimental studies, attitudes, motivation, and interest are regarded as important determinants of the quality and depth of learning processes, student persistence, and study choice (Hidi & Renninger, 2006; Maltese & Tai, 2011; Pintrich, Marx, & Boyle, 1993; Vansteenkiste, Lens, & Deci, 2006). Some authors even proposed that affective learning outcomes might be more important to further learning than cognitive ones (e.g., Maltese & Tai, 2011). Although the direct correlations between attitude and achievement as observed in empirical survey studies are too weak to warrant such claims (Ma & Kishor, 1997; Singh, Granville, & Dika, 2002), it is clear that affective outcomes are important, especially in the long run.

Research on the effect of teaching approaches on attitudes in science and mathematics education has a long and somewhat cumbersome history. In their seminal review about attitudes in science education, Osborne et al. (2003) conclude that: even a cursory examination of the domain reveals that one of the most prominent aspects of the literature is that 30 years of research into this topic has been bedeviled by a lack of clarity about the concept under investigation. (p. 1053).

This lack of clarity is reflected in the common practice that researchers will use their own newly developed instruments to assess attitudes, rather than a well-validated existing one (Blalock et al., 2008). Moreover, with regard to intervention studies, Osborne et al. conclude that, although the science education literature may contain thousands of reports of interventions designed to change attitudes, effect sizes are marginal and most findings cannot be generalized beyond the particular study. The key problem according to Osborne et al. is that studies compare student outcomes between curricula without undertaking an analysis of the ways in which the curricula differ. Although they admit a few positive exceptions (the use of life-world contexts does seem to have consistent positive effects), their sobering conclusion is that teacher variables are more important anyway, and that research may better turn away from curriculum variables.

Nevertheless, over the last ten years, several review studies have documented the effects of innovative teaching approaches on attitude and/or achievement. In a systematic review of 17 experimental studies assessing the effects of context-based and Science Technology and Society (STS) approaches in science education, Bennett, Lubben, and Hogarth (2007) found a positive effect on attitude, but no consistent effect on achievement. Likewise, a systematic review by Potvin and Hasni (2014b), based on 97 intervention studies, which covered a broader range of educational approaches such as inquiry-based, collaborative work, ICT-usage, and out-of-school activities, found a modest positive effect on attitude/motivation/interest (which were taken together in their analyses). In both reviews, all kinds of study designs were included, and there is little information about the size of the effects. With regard to achievement, meta-analyses have been conducted by Schroeder, Scott, Tolson, Huang, and Lee (2007), and by Furtak, Seidel, Iverson, and Briggs (2012). Schroeder et al. included 61 studies covering eight types of teaching innovations. They found a substantial overall effect on achievement ($d = 0.67 \pm 0.01$), with no significant differences between the approaches. Furtak et al. specifically focused on the achievement gains for inquiry-based education. They included 37 studies, and found a substantial positive effect on achievement ($d = 0.50 \pm 0.01$), with markedly higher values for teacher-led inquiry contrasted to student-led inquiry. All of the above reviews had their focus on science education. The evidence in mathematics seems to be sparser, with the available review studies focusing on other influencing factors such as teacher behavior or classroom atmosphere (e.g., Middleton & Spanias, 1999). A notable exception is Rosen (2009), who report substantially positive effect-sizes for 'problem posing', both on achievement and on attitudes towards mathematics. To sum up, although the general tendency across studies seems to be that innovative teaching approaches do have positive effects, there is little clarity about what interventions cause effects on what outcome, and under what conditions.

1.1. Affective outcomes: attitude, interest and motivation

In educational research, the influence of affective factors on learning is being addressed from at least three largely separate research traditions: attitude, interest, and motivation research. Several authors have attempted to bring unity by synthesizing perspectives (e.g., de Brabander & Martens, 2014; Eccles & Wigfield, 2002; McLeod, 1992; Van Aalderen Smeets, Walma van der Moolen, & Asma, 2012), but thus far this has not led to a widely used common framework. Yet, in order to compare the effectiveness of interventions, and to synthesize findings across studies as we do in our review, we need a conceptual framework to decide whether two measurements map on the same underlying construct, or whether they measure different constructs. Some authors have opted for rather fine-grained frameworks (e.g., Gardner, 1995), which allows for highly specific

conclusions, but limits the opportunities to compare outcomes across a wider sample of studies; others took together all interest, motivation, and attitude measures in a single construct (e.g., Potvin & Hasni, 2014b), leading to a larger pool of studies but fuzzier interpretations. In order to find an optimum between these extremes, we briefly review the underlying psychological theories.

According to commonly held definitions in psychology research, an attitude is “a summary evaluation of a psychological object captured in such attribute dimensions as good-bad, harmful-beneficial, pleasant-unpleasant, and likable-dislikable” (Ajzen, 2001, p. 28), or alternatively “a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (Eagly & Chaiken, 1993, p. 1). It should be noted that in science education, the term attitude has also been used to refer to *scientific attitudes*, such as a critical stance, or a demand for verification (e.g., Osborne et al., 2003). Although the attainment of scientific attitudes is an equally important goal for science education, they fall beyond the scope of the current review.

The field of attitude research has long been plagued by the fact that attitudes would only be poor predictors of actual behavior, even up to the point that researchers were calling to abandon the concept of attitude (e.g., Wicker, 1969). This lack of predictive power has been explained from the observations that a) attitude is a multidimensional construct (e.g., Gardner, 1975) and b) any concrete situation involves many particularities that are not being accounted for by the overall attitude assessment (e.g., Ajzen, 1991, p. 185), and c) that in addition to the attitude toward the object or behavior, behavior is also influenced by subjective norms and perceived behavioral control (Ajzen, 1985).

Subjective norms refer to the opinions significant others are supposed to hold about the behavior, and whether one would like to be seen as someone who engages in this particular behavior. With regard to the science and mathematics domains: if science is regarded as “for nerds only”, this will turn off students who do not want to be regarded as nerds. In science education research, this has been operationalized in a scale assessing the perceived Normality of Scientists.

Perceived behavioral control, or self-efficacy, is an expectancy about one's capabilities to learn or perform a given task (Bandura, 1977). In science and mathematics education research, similar constructs are variously being addressed by names such as Science Self Concept or Math Anxiety (reverse coding). Unlike measures for general self-esteem, self-efficacy toward a specific behavior has been a strong predictor of actual behavioral choices, and perseverance. These elements have been incorporated in Ajzen's *theory of planned behavior* (1991) which became a successful model for predicting human behaviors in many areas.

Like attitude theories, interest theories primarily focus on person-object relations. According to a widely held definition by Hidi and Renninger (2006), interest refers to “the psychological state of engaging or the predisposition to reengage with particular classes of objects, events, or ideas over time” (p. 112). This definition encompasses short-term situational interest (state) as well as long-term personal interest (disposition). In the present context our focus will be on personal interest as an educational outcome. As with attitude, the general developmental pattern for interest seems to be a “downward trend ... plateaued in later years, with high variability in mean levels, but little variability in the shape of the growth trajectories” (Frenzel et al. 2010, p. 507). However, as can be observed in young children as well as in scientists, the onset of a personal interest can be a quite idiosyncratic event, and the object of interest can be very narrow (e.g., American Scientist, 2012; Chi & Koeske, 1983; DeLoache, Simcock, & Macari, 2007).

Motivational theories also attempt to answer the question about what gets an individual moving (energization) and toward what activities or tasks (direction; Pintrich & Schunk, 2002). Compared to attitude and interest research, many motivation theories are more focused on external factors that influence energization for the task at hand (rewards, punishments), and other constructs are domain overarching personal traits (mastery goal orientation versus performance goal orientation). These are beyond the scope of the current review.

Other motivation theories take a more long-term perspective on what guides the direction of motivation (goal orientations, self-determination theory). A leading theory in this field is the Self Determination Theory (SDT, Ryan & Deci, 2000), which identifies the *basic psychological needs* for Competence, Autonomy, Relatedness as ultimate drivers of behavioral choices. People will tend to engage in behaviors that are likely to help satisfy these needs, and avoid behaviors that threaten the fulfillment of these needs (failure avoidance). SDT also offers a developmental perspective: Motivation can develop from externally driven to autonomous and identified for activities that successfully contribute (competence) toward self-selected (autonomy) and meaningful (relatedness) goals. SDT proposes a continuum from pure externally driven behavior (avoid punishment) to autonomous internally driven behavior. Such autonomous behavior could be driven by intrinsic motivation (just for the joy of it, which is very similar to interest, cf. Eccles & Wigfield, 2002), or it could contribute to other goals valued by the actor (becoming an engineer, being valued by the community). Both intrinsic motivation and the career perspective resemble forms of personal interest that are regularly being assessed in science and mathematics education research. The desire to be valued by the community has close relation to the concept of subjective norm in Ajzen's (1991) theory of planned behavior.

1.2. Multiple grain sizes – toward a conceptual framework

One proposed remedy for the poor predictive validity of attitudes is the aggregation of specific behaviors across occasions. By aggregating different behaviors, observed on different occasions and in different situations, other sources of influence tend to cancel each other out, with the result that the aggregate represents a more valid measure of the underlying behavioral disposition, and a more powerful predictor of general behavior (Ajzen & Fishbein, 2005; Crano & Prislín, 2005; Kind et al.,

2007). By contrast, if the aim is to predict a specific behavior, it is important to assess attitudes specific to this particular behavior, rather than attitudes toward a broad behavioral category (Ajzen, 1991).

For attitudes toward science and mathematics, this would imply that doing science in school, doing practicals, doing science as a leisure activity, choosing a science career, and support for science as a societal phenomenon should all be regarded as different attitude objects. On the other hand, Kind et al. (2007), in factor analyses on questionnaire data from 932 middle school students, found that questions about “learning science in school”, “learning science outside school”, and “future participation in science” would all load on a single factor, which they termed *general interest in science*. Likewise, some authors point to the difference between the cognitive (beliefs), affective (emotional responses) and conative (behavioral intentions) components of attitude. Even if students agree that science is important, this opinion may remain unrelated to students’ feelings about doing science themselves. Evidence is mixed as to whether these components should be accounted separately in order to predict behavior. For instance, Francis and Greer (1999) claimed a one-dimensional structure for their attitude toward science instrument, but Walker, Smith, and Hamidova (2013) using the same instrument with a new sample of students, found a two-factor structure, with the items neatly split up between beliefs about the relevance of science, and interest in doing science oneself (affective and conative).

Given the great diversity in conceptual frameworks, the broadly targeted nature of many educational interventions, and the diffuseness of the targeted behavior, we consider that both the fine grained and the more global perspective may have their merits, and we cannot justify a single level of analysis. Therefore, in this study we opt for a hierarchical framework (Fig. 1).

In line with the theory of planned behavior and attitude research, and further informed by an initial sample of about 50 studies, we identified two clusters of cognitive beliefs about the relevance of science or mathematics, namely *personal relevance* and *societal relevance*. Typical items would be “science helps me in dealing with everyday problems” and “science contributes to our wealth”, respectively. Some studies combine both kinds of items in a single scale, in which case they could be said to assess *general relevance*. Likewise, *interest* (including intrinsic motivation) was being assessed in three different contexts, namely *classroom*, *leisure*, and *career*. The three constructs taken together could be said to represent *general interest* (cf., Kind et al., 2007). Furthermore, several studies assess *self-efficacy* toward the specific domain at hand and/or the *normality of scientists*, which is taken as indicative of subjective norms/relatedness. Finally, a group of studies either combine items from different categories in a single overall measure, or employ a more fine-grained instrument while only reporting an overall score. In line with the idea that an overall attitude would predict overall behavior, these studies could be regarded as assessing the overall attitude.

1.3. Achievement

While affective outcomes are important, they should not go at the cost of the cognitive learning outcomes. Achievement may seem more straightforward to conceptualize, but achievement outcomes pose their own challenges. Most notably, and in science even more than in some other school subjects, achievement tests are topic and course-specific, and the majority of studies use achievement measures specifically constructed for the task at hand. As a consequence, the quality of instruments is hard to assess, and in many studies, the experimenter-made measures are more aligned with the content taught in the experimental condition than with the content in the control condition (e.g., inquiry skills are needed on the test, or the questions are posed in context-rich format). On the other hand, one may argue with some ground that standardized tests are often more aligned with the outcomes of the traditional curriculum. Slavin and Madden (2011), focusing on mathematics and reading studies reviewed in the U.S. Department of Education’s What Works Clearinghouse (WWC), found that measures that are “inherent” to the treatment (covering content not taught in the control group) are associated with effect sizes that are much higher compared to measures of the curriculum taught in experimental as well as control groups ($d = 0.45$ vs. $d = -0.03$). By contrast, Schroeder et al. (2007) in their meta-analysis, where 47 out of 62 achievement tests were locally constructed, found no difference in outcome ($d = 0.73$ and $d = 0.75$ resp.).

1.4. Teaching approaches

Over the past decades many teaching approaches have been proposed to foster positive attitudes toward science and mathematics. Although each approach has its own unique features, which could be characterized on many dimensions, a few

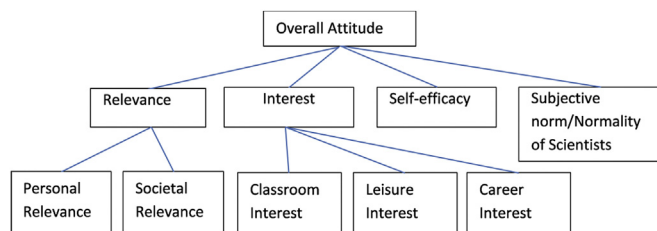


Fig. 1. Hierarchical structure of attitude constructs.

broad categories tend to be distinguished in research as well as in public debate (e.g., Potvin & Hasni, 2014b; Schroeder et al., 2007). To categorize the studies in their sample Schroeder et al. drew on a typology of teaching strategies by Wise (1996). Taking this typology as a starting point we encountered the following types of interventions in our study:

- Context-based teaching (con). Many curriculum reforms have focused on the use of contexts and applications of science and mathematics. Although the intensity and the role of context use vary across implementations, one aim of all context-based curricula is that students will experience the relevance and applicability of the science content in society and in their personal life worlds (Gilbert, 2006). Many studies on context-based interventions report gains in students' attitudes to science and technology, with learning gains similar to those of conventional approaches (cf., Bennett et al., 2007);
- Inquiry-based learning (ibl). In IBL, students get involved in inquiry activities in order to find answers to learning questions. Implementations differ in the amount of teacher guidance, but students get (partial) responsibility for devising the research questions and methods, and interpreting the results. Thus, inquiry-based learning is supposed to promote engagement and ownership, and a more "human" view of science as knowledge in the making. Many studies on inquiry-based curricula report positive effects on attitudes as well as achievement (Furtak et al., 2012; Gibson & Chase, 2002);
- ICT-rich learning environments (ict). ICT-rich teaching approaches include (individualized) computer-based instruction, games, feedback, interactive quizzes, computer based labs, simulations and robotics. It would be hard to propose a common mechanism by which ICT usage would lead to more positive attitudes, but many studies on ICT-based interventions report such gains (Potvin & Hasni, 2014b). Proposed mechanisms include that students enjoy working with computers, students feel more safe to experiment and make mistakes, and/or students appreciate the (quick) feedback;
- Collaborative learning (coll). Collaborative and cooperative teaching approaches, such as project-based work, discussion, "jigsaw" or peer feedback, tend to enhance social interaction and relatedness between learners, and quite often involve an increased ownership of the content to be learnt. Literature reports positive effects on students motivation, increased self-confidence and satisfaction, and more positive attitudes toward the subject matter (Lazarowitz & Hertz-Lazarowitz, 1998);
- Extracurricular (ext). Extracurricular activities are not part of the standard lesson plan or classroom environment, yet are part of, or are strongly linked to the school program. Examples include field trips, mobile science labs, summer camps, guest lectures, and visits to science centers. These types of interventions usually aim at raising curiosity (interest), encounters with role models/scientists (normality) and relatedness. Literature suggests that extracurricular activities can provide learners valuable and particularly motivational opportunities to learn science (Rennie & McClafferty, 1996).

It might be worthwhile to notice the categories of teaching approaches (according to the Wise/Schroeder typology) that remained empty in our research. In our final set, we did not find any research on Questioning Strategies, Focusing Strategies, Hands-on Strategies, Assessment Strategies, or Direct Instruction. It seems that the affective effects of these more cognitively oriented approaches received little research attention so far in research on attitudes.

Beyond the type of educational approach, many other factors will influence the affective outcomes of education (e.g., Myers & Fouts, 1992; Middleton & Spanias, 1999). It is desirable to include such influencing factors in the analyses, both because it can be informative to study their effects, and because inclusion of such factors can enhance the power of analyses by accounting for what otherwise would be unexplained noise. However, many relevant background characteristics are not routinely reported. With respect to the sample of participants, we expected that grade level and age, SES and gender could affect the outcomes. With respect to the intervention, we expect that more extensive interventions, and better prepared teachers (Slavin, Lake, Hanley, & Thurston, 2014) will have stronger effects. Furthermore, although our general question is about attitudes in science and mathematics, there are significant differences in popularity between biology and for example physics (cf., Osborne et al., 2003), and it was expected that the effectiveness of an educational approach may depend on the domain at hand.

A possible caveat when talking about innovative teaching approaches is that they are always innovative relative to a, mostly implied, "regular" practice, and this regular practice can vary across time and space. Context-based teaching, for instance, has become more or less mainstream in many countries nowadays, but not in others. To be included in our analyses, studies had to have a complete comparative design, so the teaching intervention is well-defined as the difference between the intervention and the control conditions. However, when synthesizing across studies, the definition of "regular" teaching might differ across studies and, given a lack of information and a limited number of studies, there is no way to control for that. This implies that our analysis has to rely on a linear additive effects model, ignoring the fact that innovations might work out differently for different regular teaching practices.

The purpose of the current review is to find out how the respective innovative teaching approaches in science and mathematics education do affect (components of) attitude and achievement as defined in the previous sections. Specific questions are:

- 1) What is the impact of innovative teaching approaches on student attitudes?
- 2) What is the impact of innovative teaching approaches on student achievement?
- 3) Are the answers to a and b different dependent on duration of the intervention, grade level, preparedness of the teachers, type of innovative approach, or school subject?
- 4) Is there a tradeoff between promoting positive attitudes, and promoting achievement?

2. Methods

2.1. Literature search process

The literature on attitudinal outcomes of science and mathematics education was searched broadly, canvassing the Web of Science, ERIC, PsycINFO, and Scopus databases from 1988 to 2014. As a quality criterion, we decided to limit our search to peer reviewed literature. In a stepwise process we defined a query consisting of Attitude \times Domain \times Design keywords (for the full query, see the [Methods supplement accompanying the online article](#)). Since the initial query yielded over 15,000 publications, we refined the query by adding filters to limit the number of search results without excluding any usable studies. The first 100 hits in each database were used to verify that these additional filters did not result in loss of relevant publications. After deduplication, 6066 unique publications remained. The search in Scopus and the Web of Science yielded many more studies than in ERIC and PsycINFO, presumably because many studies on interest have been published in science and mathematics education journals, rather than educational or psychological journals (see [Fig. 2](#)).

2.2. Classification and selection

The classification and selection process was facilitated by organizing all publications, meta data and coding information in a database, such that we could keep track of each publication throughout the entire process. All authors participated in the coding team.

Step 1: Surface level screening inclusion and exclusion criteria. The purpose of the first step was to efficiently exclude a substantial part of studies that did not meet our inclusion criteria. This surface level screening was mostly done on the basis of reading title and abstracts. Only in the few cases where essential study characteristics were not evident from title and abstract, the full text was consulted. Inclusion criteria were (for coding guidelines see the [Methods supplement accompanying the online article](#)):

1. Curriculum-related intervention in science (physics, chemistry, biology) or mathematics education;
2. Participants are students in general primary or secondary education (Grades 1–12 in the US system);
3. Domain-specific attitude, interest, or motivation were assessed quantitatively;
4. The study design was experimental or quasi-experimental (control group and pre-posttest design).

After this screening 533 publications remained.

Step 2: Design check and categorization. In this step, full texts were consulted to verify that all statistical information required for inclusion in a meta-analysis was reported in the publication. For the purpose of the present report, only studies with an experimental or quasi experimental design were included, and they had to report means, standard deviations, and sample sizes for both treatment and control, and for both pre- and posttest, or equivalent information. After this step 152 publications remained.

Step 3: Quality of attitude measurement and categorization of interventions. This step involves a more in-depth reading of the full text to assess the quality and type of the measurement(s) and the type of intervention. In order to be included, an attitude measurement scale had to assess domain-specific attitudes (e.g., “learning physics is fun, rather than “learning is fun”, or “this lesson was fun”), and the quality of the instrument had to be validated in a previous publication, or reliability statistics and at least three sample items per scale had to be reported in the publication at hand. If a publication had at least one acceptable attitude scale, both the attitude and the intervention were categorized according to our framework,

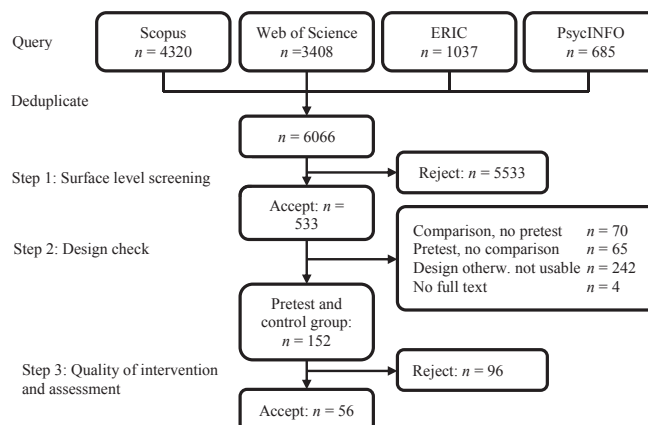


Fig. 2. Flow diagram of publication selection process.

and statistical data were extracted. In some cases it turned out that the reported information did not allow for the computation of effect sizes after all. If the study reported achievement outcomes, these outcomes were also extracted. Moreover, study level moderating variables were extracted in the same pass. Unfortunately, many relevant moderating variables are not regularly reported, and information that is reported can be in many incompatible formats. The moderating variables that were reported sufficiently often to be included in our further analyses were: domain (biology, chemistry, physics, mathematics, or general science); time interval (days from first to last activity); whether the teachers had been trained on the approach; and grade level of the students. Since grade-level systems differ across countries, grade levels for non-US studies were converted to the grade levels for the corresponding age group in the US-system (i.e., grade 1 starts at the age of 6).

In 11 eligible publications, multiple (independent) experimental conditions were compared. For eight of these (Alexander, Fives, Buehl, & Mulhern, 2002; Cebesoy & Akinoglu, 2012; Eskander, Bayrami, Vahedi, & Ansar, 2013; Hong & Lin-Siegler, 2011; Isiksal & Askar, 2005; Kara & Yeşilyurt, 2008; Ke & Grabowski, 2007; Tarim & Akdeniz, 2008) the different treatment conditions were assigned to the same intervention type in our analysis. Therefore, data for the multiple conditions were pooled into a combined estimate of the pretest and posttest means and standard deviations (Higgins & Green, 2011, Chapter 7). One study (Hwang, Shi, & Chu, 2011a) had one experimental condition and two control groups. However, the experimental and one of the control groups were both exposed to the same educational approach, although in a slightly different manner. Therefore, we pooled the data of the experimental group with one of the control groups. In two publications the multiple treatment conditions were so different that we analyzed them separately (Cinici, Sozibilir, & Demir, 2011; Peşman and Özdemir, 2012). In these cases, the control group was split in two, and each experimental condition was paired with half the control group data. In three publications (Barnett et al., 2004; Akçay, Yager, Iskander, & Turgur, 2010; Rosen, 2009) an experiment was done twice with different groups of participants, for instance, one in elementary and one in secondary school. We included these experiments separately in our analysis. Finally, the publication by Chiu (2011) reported on girls and boys separately. Since our research question does not make this distinction we included the combined results from boys and girls in our analyses.

After this step, 56 publications with 65 independent experiments were left for further analysis. Table 1 presents an overview of experiments by domain and intervention types. The interventions labelled “Other” could be further subdivided into meta cognitive learning strategies (4×), use of representational formats (3×), and conceptual change approach (1×). (For an overview of individual experiments with study characteristics, see the [Methods supplement accompanying the online article](#)).

2.3. Quality of the coding process

In order to ensure the quality of the coding process, the team regularly met up to discuss the interpretation of the code book rules for problematic cases. Consistency of the coding in step 2 and 3 was verified by having one of the authors, who had not been involved in the initial coding for those steps, recode a randomly selected set of about 10% of publications in that step. Interrater agreement, as expressed in Cohen's Kappa (Cohen, 1968) was found acceptable (Table 2).

2.4. Statistical analysis

For each study the effect size per outcome variable was calculated as the standardized difference in mean change for the intervention and control condition,

$$d_k = C_p \left[\frac{(M_{post,T} - M_{pre,T}) - (M_{post,C} - M_{pre,C})}{SD_{pre}} \right] \quad (1)$$

with M being the mean at pretest (pre) or posttest (post) for the treated (T) or control group (C). SD_{pre} is the pooled pretest standard deviation for the intervention and control condition, and C_p the correction factor for small sample sizes (Morris, 2008). Standard errors for the effect sizes were calculated as proposed by Morris (2008). We ran separate random effect meta-analyses per outcome variable using inverse variance weighting to account for differences in sample size across studies.

Table 1
Experiments included by domain and intervention type.

Domain	#Experiments	Intervention type					
		Inquiry	Context	ICT	Collaborative	Extracurr.	Other
Science, general	14	0	8	2	3	0	1
Mathematics	19	1	3	11	5	0	3
Physics	11	7	3	2	1	0	0
Chemistry	8	1	1	3	0	1	2
Biology	13	3	5	1	3	3	2
Total	65	12	20	19	12	4	8

Note: totals do not add up since one intervention can have multiple labels.

Table 2
Interrater reliabilities for steps in the analysis procedure.

Step	#	Cohen's K
2: Assignment to design categories (publication level)	39	0.64
3a: Measurement check (publication level)	14	0.65
3b: Intervention classification		
- One code of one or more	16	0.77
- Only single coding	12	0.69
3c: Assignment to attitude constructs	26	0.76

All calculations for effect sizes and standard errors were carried out in R 3.1.0 (code available upon request) and all meta-analyses were done with the Metafor package (Viechtbauer, 2010).

In order to detect whether publication bias had skewed our sample, we looked at the variation in study outcomes as a function of sample size/standard errors. In an unbiased sample, studies with a large error are expected to have more varied outcomes than studies with a small error, but the mean outcomes should not be systematically different for small and large error studies. If accidental negative outcomes remain unpublished, or accidental positive outcomes get published, this would be expected to occur more frequently for the large error studies, which in turn would lead to systematically more positive mean outcomes for large error studies than for small error studies. In case a significant meta-analytic result was found for a specific outcome variable in a set of at least ten studies, we checked whether this kind of distortion had occurred using funnel plots and regression tests from the Metafor package in R (Egger, Smith, Schneider, & Minder, 1997; Viechtbauer, 2010). (For additional details about the analysis, see the [Methods supplement accompanying the online article](#)).

3. Results

We will first report the overall meta-analyses for the effect of innovative teaching approaches on overall attitude and achievement. Next, we will zoom in on the specific attitude components, and on differential effects for specific types of teaching approaches.

3.1. Overall effects of innovative approaches on overall attitude and achievement

For 65 independent experiments the effects of an intervention on overall attitude were examined. From these, 39 experiments just reported a single outcome for “overall attitude”, another 26 experiments reported separate outcomes for specific subscales, which were combined into a single outcome per experiment for the present analysis. The forest plot in Fig. 3 shows the estimated effect size per experiment and the overall effect size and 95% CI over the 65 experiments. The results show an overall significant and large treatment effect, with $d = 0.58$ [0.37; 0.79], and $p < 0.0001$. However, the sample exhibited substantial heterogeneity, $Q(df = 64) = 200.83$, $p < 0.0001$. The experiment by Prokop, Tuncer, and Kvasničák (2007) and the four experiments by Akçay et al. (2010) strongly deviate from the general pattern. After removal of these studies, the effect remains significant $d = 0.35$ [0.24; 0.47], $p < 0.0001$, and no significant heterogeneity remains ($Q(df = 59) = 54.78$, $p = 0.63$).

With regard to achievement, 42 experiments out of 65 measured achievement in addition to overall attitude. Since content and learning aims differ per study, there is no standard measurement instrument, and all we can conclude about when synthesizing across studies is a “general” effect on achievement. For the 42 experiments, we found a significant and large effect on achievement, $d = 1.07$ [0.68; 1.47], $p < 0.001$ (Fig. 4). However, the sample exhibited substantial heterogeneity, $Q(df = 41) = 199.29$, $p < 0.0001$. Two studies (Acar Sesen & Tarhan, 2013; Prokop et al., 2007) were considered as outliers and removed from the statistical analysis. Without these two studies, the effect remained significant and substantial ($d = 0.78$ [0.60; 0.97], $p < 0.001$), and some heterogeneity was left, $Q(df = 39) = 59.15$, $p = 0.02$. The latter implies that the sample of studies thus cannot be regarded as taken from a single normally distributed population of studies; part of the variance is likely to be accounted for by differences in underlying study characteristics. However, at $I^2 = 34\%$, the heterogeneity is small enough to regard the overall effect as a meaningful outcome (Higgins & Green, 2011).

Next, we verified whether there was an association between overall attitude effect and achievement effect at the study level. The correlation between the two was found nonsignificant, $r = 0.26$, $N = 42$, $p = 0.091$. After removal of the outliers (Acar Sesen & Tarhan, 2013; Prokop et al., 2007), the outcome remains nonsignificant $r = 0.17$, $N = 40$, $p = 0.30$.

3.2. Effects of type of approach, domain, and background variables

To further assess causes behind these effects we performed a meta-regression with type of teaching approach, domain, timeframe (days from start to end of the intervention), grade level, and the presence of teacher training as predictor variables. The variance in each of the predictor variables was sufficient to make the analysis meaningful (timeframe: $M = 88$ days, $SD = 125$; grade level: $M = 7.4$; $SD = 2.2$; and 22 out of 65 studies had some form of teacher training). Information on grade level was missing for three studies (Acar Sesen & Tarhan, 2013; Hwang et al., 2011a; Uzuntiryaki and Geban, 2005), we

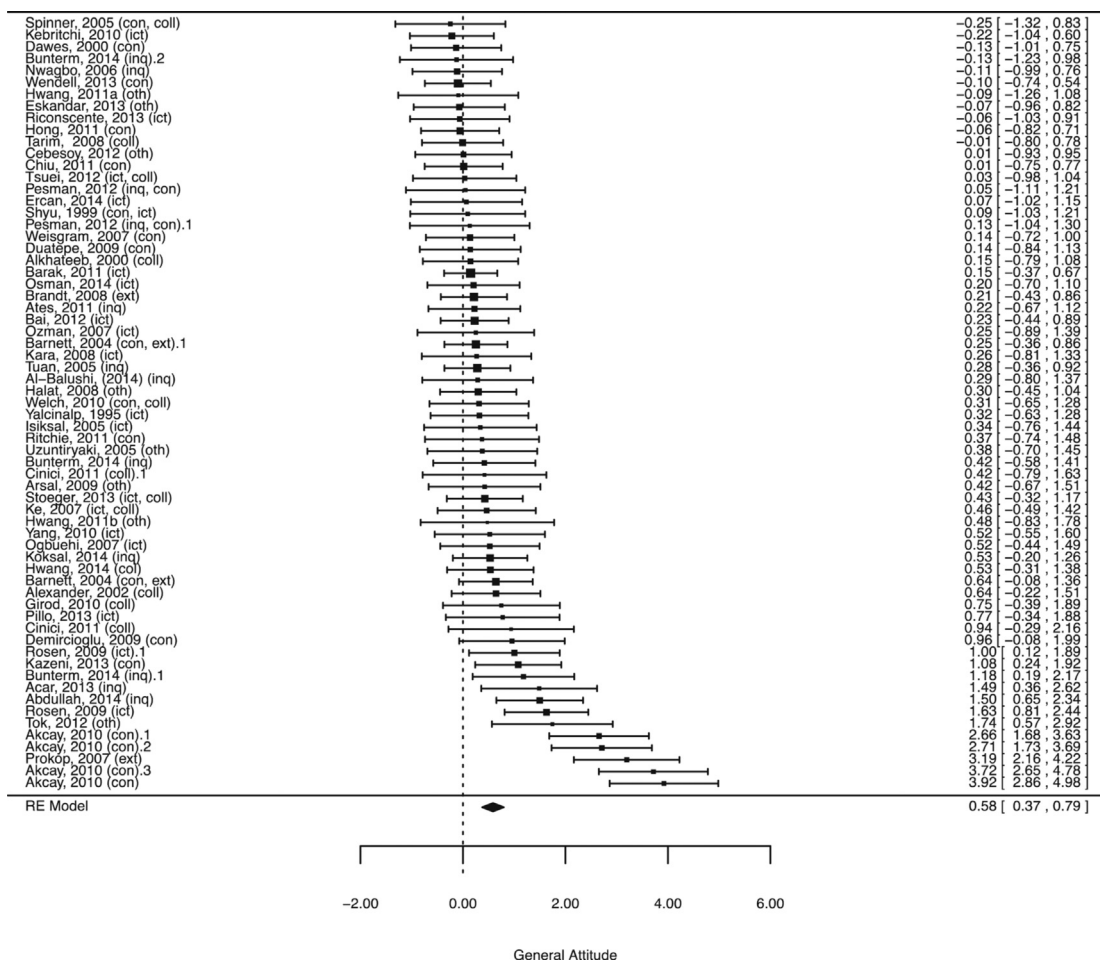


Fig. 3. Effect sizes for overall attitude (including aggregated subscales). Studies are labelled with their first author, and intervention types are indicated between brackets.

imputed the average grade level of the other studies. For the five studies (which comprised six experiments) that did not report whether teacher training had taken place (Ke and Grabowski, 2007; Pilli & Aksu, 2013; Rosen, 2009; Uzuntiryaki and Geban, 2005; Yang and Tasi, 2010) we assumed absence of teacher training.

In Table 3 the result of the meta-regression analysis is presented for the set of studies excluding the experiments by Akçay et al. (2010). The data provide no evidence for one approach being more effective than the others, no evidence for differences between domains, and also teacher training and timeframe do not explain significant amounts of variance. The negative effect of grade level is significant however, indicating that the effectiveness of interventions decreases for older students. A similar meta-regression was conducted for achievement but no significant effects were found.

3.3. Effects on specific attitude constructs

Next, we analyzed subsets of studies separately per subscale. There were no studies to report on personal relevance as a separate construct. For all other studies an effect size could be computed. In all cases the data from Prokop et al (2007) were removed as outliers. The resulting effect-sizes are presented in Table 4 (forest plots are presented in the Results supplement accompanying the online article). The effects for General Interest and Career Interest are significant and positive. However, there are no significant differences between any of the subscales, so there is no ground to declare one of the effects larger than the others.

3.4. Publication biases

The findings in a meta-analysis might be distorted by publication bias. For each significant result obtained in the previous analyses we used funnel plots and regression analysis to assess whether the distribution of outcomes suggested a publication

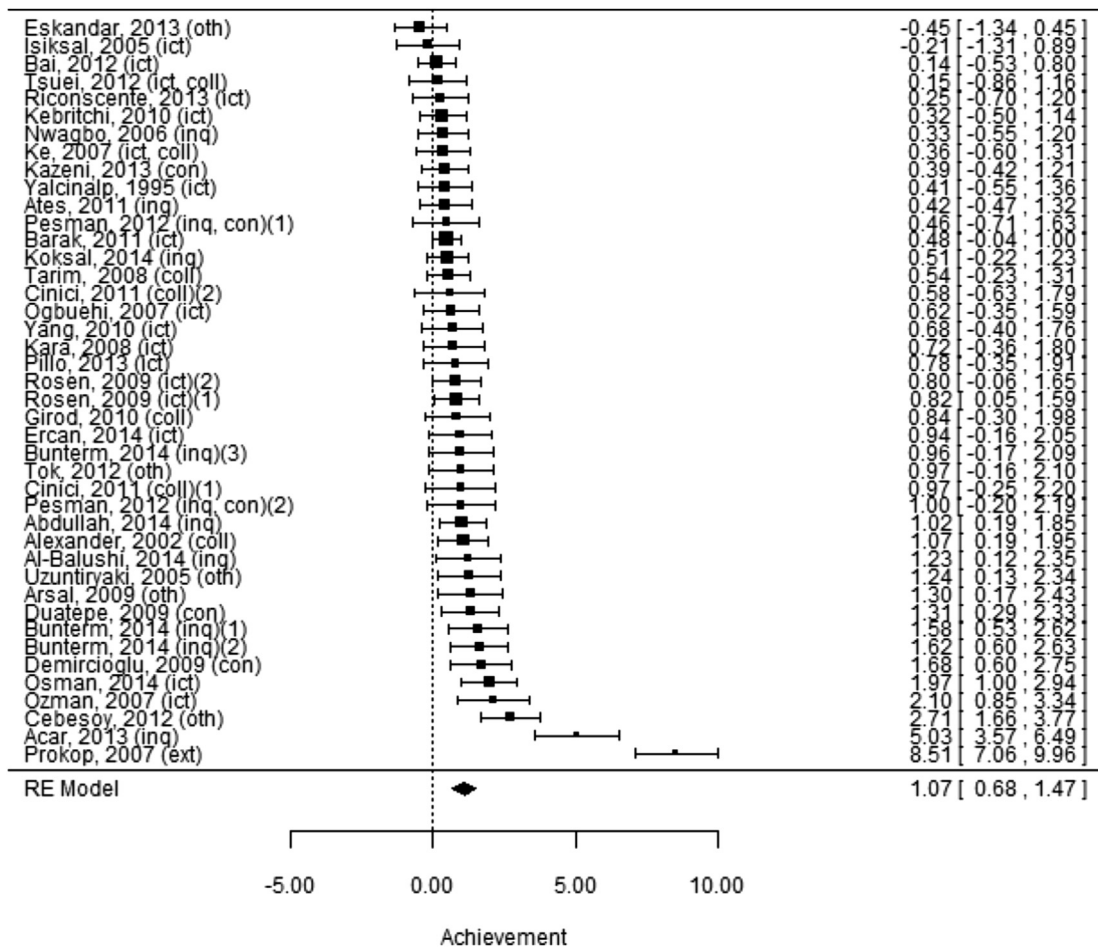


Fig. 4. Effect sizes for achievement.

bias (plots are presented in the [Results supplement accompanying the online article](#)). After removal of the outliers we had already identified in our previous analyses, the remaining results are well in line with the expected funnel shape, and all regression tests were nonsignificant. From these analyses we find no indications for publication bias.

3.5. Conclusion and discussion

The purpose of this study was to describe the effects of innovative teaching approaches on student attitudes and achievement in science and mathematics education. Based on previous reviews, and the interventions we found in the current sample of studies, we distinguished five types of commonly advocated educational approaches: context-based, inquiry-based, ICT-enriched, collaborative, and extra-curricular. From a theoretical stance, we consider attitude a multidimensional construct comprising: perceived relevance (personal and societal), interest (school, leisure, career), self-efficacy, and normality of scientists. However, many experimental studies report a global aggregate measure rather than scores for each specific subscale, and it would imply a substantial loss of information if we would exclude all these studies. Moreover, an “overall attitude” construct may have its own merits since study behavior and study choice are influenced by a large number of different attitudes. Therefore, in this review we adopted a hierarchical framework to allow both fine-grained and global analyses.

In answer to the first research question, about the effects of innovative teaching approaches on attitudes, we found significant positive effects on Overall Attitude (61 studies, $d = 0.35$ [0.24; 0.47]), General Interest (20 studies, $d = 0.22$ [0.04; 0.40]), and Career Interest (4 studies, $d = 0.40$ [0.04; 0.76]). These values are well in line with the typical value of 0.2 for affective outcomes as reported by Hattie (2009). For the other attitude components uncertainties were large, which might be partially explained by the small number of studies per specific attitude construct.

In answer to the second research question, about the effects of innovative teaching approaches on achievement, we found a significant and large effect ($d = 0.78$ [0.60; 0.97]). This value is considerably larger than the effects for attitude, and well

Table 3

Meta-regression of Overall Attitude on Moderating Variables (with Akçay et al. excluded).

	B	SE	95% CI		p
Intercept	0.91	0.33	0.26	1.56	0.01**
Timeframe	–0.0009	0.0006	–0.002	0.0002	0.10
Grade level	–0.10	0.04	–0.17	–0.02	0.01*
Teacher training	0.05	0.18	–0.29	0.40	0.76
Domain: mathematics	–0.07	0.21	–0.48	0.33	0.72
Domain: physics	0.33	0.25	–0.17	0.82	0.20
Domain: chemistry	0.18	0.27	–0.35	0.72	0.51
Domain: biology	0.31	0.24	–0.15	0.78	0.19
Int: Inquiry	0.13	0.23	–0.33	0.58	0.59
Int: Context	–0.02	0.19	–0.40	0.35	0.90
Int: ICT	0.17	0.18	–0.18	0.53	0.34
Int: Collaborative	0.10	0.19	–0.28	0.48	0.60
Int: Extracurricular	0.24	0.27	–0.29	0.76	0.38

Note. B: Regression coefficient, SE: standard error, CI: 95% confidence interval. * $p < 0.05$ ** $p < 0.001$.

Table 4

Meta-analytic result per subscale.

Subscale	#Studies	Effect Size (<i>d</i>)	SE	95% CI		Test of heterogeneity		
				Lower	Upper	Q	df	p
General relevance	4	0.15	0.18	–0.19	0.50	0.44	3	0.931
Personal	0	–	–	–	–	–	–	–
Societal	3	0.22	0.21	–0.18	0.63	0.12	2	0.940
General interest	20	0.22*	0.09	0.04	0.40	1.51	19	0.906
Classroom	4	0.16	0.18	–0.19	0.51	0.90	3	0.825
Leisure	2	0.08	0.23	–0.38	0.53	0.36	1	0.548
Career	4	0.40*	0.18	0.04	0.76	1.91	3	0.591
Self-efficacy	14	0.12	0.11	–0.09	0.33	2.95	13	0.998
Normality of scientist	4	0.16	0.24	–0.31	0.62	2.02	3	0.578

Note. Numbers of studies excluding outliers, effect size (*d*) and standard error (SE), 95% Confidence Intervals (CI), heterogeneity statistic Q with degrees of freedom (df) and *p*-value. * $p < 0.05$.

above the typical value of 0.4 for achievement as reported by Hattie (2009). However, in interpreting the effect size for achievement, it should be taken into account that achievement measures were not subjected to the same rigorous criteria we applied to the attitude scales. Standardized achievement tests are rare in this corpus of studies, so each study has its own achievement test specifically designed for the content at hand, and it might be that in some cases the achievement test was biased toward the innovative group (see Slavin & Madden, 2011).

To answer the third research question about the differential effects of different types of interventions, we conducted a meta-regression with type of teaching approach, duration, teacher preparation and domain as independent variables. The overall attitude construct was the only one with a sufficient number of studies to perform this analysis. We found a significant negative effect for grade level, which implies that the attitudes of older children are harder to influence. There were no significant effects for type of teaching approach, duration, teacher training, or domain.

To answer the final research question about a tradeoff between promoting attitude and promoting achievement, we tested whether studies that were more successful in promoting positive attitudes would also be more successful in promoting achievement. However, no correlation was found between the effect of an intervention on attitude and its effect on achievement.

We had expected different effects for different types of teaching approach and stronger effects for longer lasting interventions and interventions with teacher training, but no such effects were found. The fact that we did not find any significant differences between approaches could be interpreted in several ways. We might conclude that any innovation is good, or that any positive effects are just novelty effects. In our view, a more likely interpretation is that, rather than type of approach as defined in this study, it is the quality of the content and the implementation that matters. A context can be dull or inspiring; productive for content learning or just superficial; an inquiry assignment can be confusing or stimulating; and a teacher can amplify the message of an approach or interfere with it. Such features might be essential for the effects of a teaching approach, and they certainly are essential for everyone who intends to replicate the experiment, or implement the proposed approach. As a further complication, the effect of a teaching innovation might be different for different baseline situations. If such interactions are strong, the entire concept of generalizability would be at stake. Our relatively homogeneous data do not point in this direction, but at least more detailed information about the teaching in the “regular” or “control” conditions would be needed. Unfortunately, with the notable exception of Girod, Twyman, and Wojcikiewicz (2010), details about the teaching in either condition were only seldomly reported in the studies we encountered.

Likewise with regard to duration and teacher preparation, many studies provide only limited information, so timeframe had to be used as a proxy for the extensiveness of the intervention, and for teacher preparation, all studies with some form of teacher preparation were coded as “yes”. However, it should be expected that teacher training needs to be substantial in order to be measurably effective in student achievement and attitude (Hofstein, Kesner, & Ben-Zvi, 1999).

Some studies reported results much more positive than the rest. They were regarded as outliers and removed from the statistical analysis. A deviant result could have many reasons, but still we regard it useful to analyze these studies more in depth to see whether the particular conditions or interventions could offer a likely explanation for deviant outcomes. The intervention by Akçay et al. (2010), conducted in Turkey, involved a comprehensive student-centered STS curriculum. They ran four parallel experiments in grades 6 to 9. The timespan of the intervention was 180 days, which is considerably longer than the average of the other studies in our sample (94 days) but not extreme (the longest intervention in the sample lasted 2 years). What makes the Akçay et al. study really stand out is the extensive teacher training program (4-week summer camp, participation in the design, several 3-day return meetings etc.).

At the other extreme of the time scale, Prokop et al. (2007), in an experiment with 6th graders in Slovakia, found strong and lasting effects on all types of interest in biology for a one-day (10 h) fieldtrip where they visited different ecosystems, captured animals and plants, identified their species and wrote their diaries. Although the time span is very short, it seems that this intensive submersion in nature left a lasting impression on many students. It may also be a special affordance of the ecology domain that it lends itself to such an approach.

Unfortunately, we had to exclude many studies because basic information was missing; instruments were used without traceable validity information or means and standard deviations were partially missing. The same holds for several moderating variables. Many authors before us have ended their studies with the wish that future researchers will choose among the few well-established attitude instruments, rather than keep coming up with new home baked instruments. Furthermore, as a minimal set of study characteristics, we would propose that studies need to report composition of the class in terms of age, gender, cultural background, and SES; grade level; educational track; time frame and actual duration of student activities; time frame and duration of teacher preparation. However, we are convinced that such statistical parameters are not enough, and not even the most important information to predict the effectiveness of educational innovations. We would like to join Furtak et al. (2008) in their call for a fuller description of educational interventions, such that the reader would be able to replicate the intervention or at least judge both the type and quality of the intervention herself.

With regard to the practice of teaching, recommendations can only be tentative. First, since younger children have more positive attitudes toward science and mathematics, and since our findings indicate that the attitudes of younger children are more malleable, it seems worthwhile to start early on with fostering and maintaining positive attitudes, rather than try to start repairing after the decline in attitude has begun. This is especially true because interest at the onset of secondary school has been found the major predictor of interest at the end of school (Sadler, Sonnert, Hazari, & Tai, 2012). Second, since we found modest positive effects on overall attitude, and positive effects on achievement, with no significant difference between approaches, we conclude that there is no evidence to withhold teachers from investing all their best efforts in implementing high quality innovative approaches that suit their learning aims and that appeal to them most.

Acknowledgments

This research was supported by the Dutch Research Foundation NWO under a grant scheme for educational literature review studies (grant PROO411-12-232). The funding agency had no influence on the set-up of the study.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.edurev.2016.07.003>.

References*

- AAAS. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- *Acar Sesen, B., & Tarhan, L. (2013). Inquiry-based laboratory activities in Electrochemistry: High school students' achievements and attitudes. *Research in Science Education*, 43, 413–415. <http://dx.doi.org/10.1007/s11165-011-9275-9>.
- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl, & J. Beckmann (Eds.), *Action-control: From cognition to behavior* (pp. 11–39). Heidelberg: Springer.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179–211.
- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, 52, 27–58.
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173–221). Lawrence Erlbaum.
- *Akçay, H., Yager, R. E., Iskander, S. M., & Turgur, H. (2010). Change in student beliefs about attitudes toward science in grades 6–9. *Asia-Pacific Forum on Science Learning and Teaching*, 11, 1–18.
- *Alexander, P. A., Fives, H., Buehl, M. M., & Mulhern, J. (2002). Teaching as persuasion. *Teaching and Teacher Education*, 18, 795–813.
- American Scientist. (2012). 100 reasons to become a scientist or engineer. *American Scientist*, 100, 300.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215.

* References marked with an asterisk indicate studies included in the meta-analysis.

- *Barnett, M., Strauss, E., Rosca, C., Langford, H., Chavez, D., Deni, L., et al. (2004). Improving urban youth's interest and engagement through field-based scientific investigations. In *Proceedings of the 6th international conference on Learning sciences* (pp. 73–80). International Society of the Learning Sciences.
- Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education*, 91, 347–370.
- Blalock, C., Lichtenstein, M. L., Owen, S., Pruski, L., Marshall, C., & Toepperwein, M. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935–2005. *International Journal of Science Education*, 30, 961–977.
- de Brabander, C. J., & Martens, R. L. (2014). Towards a unified theory of task-specific motivation. *Educational Research Review*, 11, 27–44.
- Chi, M. T., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, 19, 29–39.
- *Chiu, M. S. (2011). Effects of a women-in-sciences/men-in-humanities intervention on Taiwanese adolescents' attitudes towards learning science. *The Asia-Pacific Education Researcher*, 20, 322–335.
- *Cinici, A., Sozbilir, M., & Demir, Y. (2011). Effect of cooperative and individual learning activities on students' understanding of diffusion and osmosis. *Eurasian Journal of Educational Research*, 43, 19–36.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213–220.
- Crano, W. D., & Prislin, R. (2005). Attitudes and persuasion. *Annual Reviews of Psychology*, 57, 345–374.
- DeLoache, J. S., Simcock, G., & Macari, S. (2007). Planes, trains, automobiles – And tea sets: Extremely intense interests in very young children. *Developmental Psychology*, 43, 1579–1586.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634.
- European Commission (Ed.). (2004). *Europe needs more scientists!*. Brussels: European Commission. URL: http://ec.europa.eu/research/conferences/2004/sciprof/pdf/conference_review_en.pdf.
- Fortus, D. (2014). Attending to affect. *Journal of Research in Science Teaching*, 51(7), 821–835.
- Francis, L. J., & Greer, J. E. (1999). Measuring attitude towards science among secondary school students: The affective domain. *Research in Science and Technological Education*, 17, 219–226.
- Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, 20, 507–537.
- Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., et al. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education*, 21, 360–389.
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82, 300–329. <http://dx.doi.org/10.3102/0034654312457206>.
- Gardner, P. L. (1975). Attitudes to science: A review. *Studies in Science Education*, 2, 1–41.
- Gardner, P. L. (1995). Measuring attitudes to science. *Research in Science Education*, 25, 283–289.
- George, R. (2006). A cross-domain analysis of change in students' attitudes toward science and attitudes about the utility of science. *International Journal of Science Education*, 28, 571–589.
- 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.
- *Gilbert, J. K. (2006). On the nature of “context” in chemical education. *International journal of science education*, 28, 957–976.
- Girod, M., Twyman, T., & Wojcikiewicz, S. (2010). Teaching and learning science for transformative, aesthetic experience. *Journal of Science Teacher Education*, 21, 801–824. <http://dx.doi.org/10.1007/s10972-009-9175-2>.
- Hattie, J. (2009). *Visible learning: A synthesis over 800 meta-analyses relating to achievement*. Abingdon: Routledge.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. http://dx.doi.org/10.1207/s15326985sep4102_4.
- Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. The Cochrane collaboration. URL: <http://www.cochrane.org/handbook>.
- Hofstein, A., Kesner, M., & Ben-Zvi, R. (1999). Student perceptions of industrial chemistry classroom learning environments. *Learning Environments Research*, 2(3), 291–306.
- *Hwang, G. J., Shi, Y. R., & Chu, H. C. (2011a). A concept map approach to developing collaborative Mindtools for context-aware ubiquitous learning. *British Journal of Educational Technology*, 42, 778–789. <http://dx.doi.org/10.1111/j.1467-8535.2010.01102.x>.
- Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science*, 320, 454–455.
- *Ke, F., & Grabowski, B. (2007). Gameplaying for math learning: Cooperative or not? *British Journal of Educational Technology*, 38, 249–259. <http://dx.doi.org/10.1111/j.1467-8535.2006.00593.x>.
- Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education*, 29, 871–893.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75–86.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction effects of direct instruction and discovery learning. *Psychological Science*, 15, 661–667.
- Lacey, T. A., & Wright, B. (2009). Occupational employment projections to 2018. *Monthly Labor Review*, 132, 82–123.
- Lazarowitz, J., & Hertz-Lazarowitz, (1998). Co-operative learning in the science curriculum. In B. J. Fraser, & K. G. Tobin (Eds.), *International handbook of science education* (pp. 449–470). Dordrecht: Kluwer.
- Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, 28, 89–120.
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, 95, 877–907.
- McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). New York, NY: Macmillan Publishing Co.
- Middleton, J. A., & Spanias, P. A. (1999). Motivation for achievement in Mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education*, 30(1), 65–88. <http://doi.org/10.2307/749630>.
- Morris, S. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386.
- Myers, R. E., & Fouts, J. T. (1992). A cluster analysis of high school science classroom environments and attitude towards science. *Journal of Research in Science Teaching*, 29, 929–937.
- OECD. (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: Organisation for Economic Co-operation and Development.
- OECD. (2008). *Encouraging student interest in science and technology studies*. Paris: Organisation for Economic Co-operation and Development.
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25, 1049–1079.
- *Peşman, H., & Özdemir, Ö. F. (2012). Approach–Method Interaction: The role of teaching method on the effect of context-based approach in physics instruction. *International Journal of Science Education*, 34, 2127–2145. <http://dx.doi.org/10.1080/09500693.2012.700530>.

- *Pilli, O., & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in North Cyprus. *Computers & Education*, 62, 62–71.
- Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167–199.
- Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in Education: Theory, research and applications*. Upper Saddle River: Merrill Prentice Hall.
- Potvin, P., & Hasni, A. (2014a). Analysis of the decline in interest towards school science and technology from grades 5 through 11. *Journal of Science Education and Technology*, 1–19.
- Potvin, P., & Hasni, A. (2014b). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, 1–45.
- *Prokop, P., Tuncer, G., & Kvasničák, R. (2007). Short-term effects of field programme on students' knowledge and attitude toward biology: A slovak experience. *Journal of Science Education and Technology*, 16, 247–255. <http://dx.doi.org/10.1007/s10956-007-9044-8>.
- Rennie, L., & McClafferty, T. (1996). Science centers and science learning. *Studies in Science Education*, 27, 53–98.
- Rocard, M., Csermley, P., Jorde, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science education now: A renewed pedagogy for the future of europe*. Brussels: European Commission. URL: http://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf (retrieved 14 September 2012).
- *Rosen, Y. (2009). The effects of an animation-based on-line learning environment on transfer of knowledge and on motivation for science and technology learning. *Journal of Educational Computing Research*, 40, 451–467. <http://dx.doi.org/10.2190/EC.40.4.d>.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education*, 96, 411–427.
- Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T., & Lee, Y. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44, 1436–1460. <http://dx.doi.org/10.1002/tea.20212>.
- Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *Journal of Educational Research*, 95(6), 323–332.
- Sjøberg, S., & Schreiner, C. (2010). *The ROSE project: An overview and key findings*. Oslo: University of Oslo.
- Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching*, 51, 870–901.
- Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, 4, 370–380.
- *Uzunıtyraki, E., & Geban, Ö. (2005). Effect of conceptual change approach accompanied with concept mapping on understanding of solution concepts. *Instructional Science*, 33, 311–339. <http://dx.doi.org/10.1007/s11251-005-2812-z>.
- Van Aalderen-Smeets, S. I., Walma van der Molen, J. H., & Asma, L. J. F. (2012). Primary teachers' attitude toward science: A new theoretical framework. *Science Education*, 96, 158–182.
- Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational psychologist*, 41, 19–31.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.
- Walker, D. A., Smith, M. C., & Hamidova, N. I. (2013). A structural analysis of the attitudes toward science scale: Students' attitudes and beliefs about science as a multi-dimensional composition. *Multiple Linear Regression Viewpoints*, 39(2), 38–48.
- Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, 25, 41–78.
- Wise, K. C. (1996). Strategies for teaching science: What works? *The Clearing House*, 69, 337–338.
- WRR. (2013). *Naar een lerende economie. WRR-rapport nr. 90*. Amsterdam University Press.
- *Yang, D. C., & Tsai, Y. F. (2010). Promoting sixth graders' number sense and learning attitudes via technology-based environment. *Educational Technology & Society*, 13, 112–125.

Further reading

- *Alkhateeb, H., & Jumaa, M. (2000). Middle-school students of United Arab Emirates: Effects of heterogeneous small group work on attitudes towards mathematics. *Perceptual and Motor Skills*, 91, 483–490.
- *Arsal, Z. (2009). The impact of self-regulation instruction on mathematics achievements and attitudes of elementary school students. *Egitim ve Bilim*, 34, 152, 1–12.
- *Ates, Ö., & Eryılmaz, A. (2011). Effectiveness of hands-on and minds-on activities on students' achievement and attitudes towards physics. *Asia-Pacific Forum on Science Learning and Teaching*, 12, 1–22.
- *Bai, H., Pan, W., Hirumi, A., & Kebritchi, M. (2012). Assessing the effectiveness of a 3-D instructional game on improving mathematics achievement and motivation of middle school students. *British Journal of Educational Technology*, 43, 993–1003. <http://dx.doi.org/10.1111/j.1467-8535.2011.01269.x>.
- *Barak, M., Ashkar, T., & Dori, Y. J. (2011). Learning science via animated movies: Its effect on students' thinking and motivation. *Computers & Education*, 56, 839–846.
- *Brandt, A., Möller, J., & Kohse-Höinghaus, K. (2008). Was bewirken außerschulische Experimentierlabors? Ein Kontrollgruppenexperiment mit Follow up Erhebung zu Effekten auf Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie*, 22, 5–12.
- *Cebesoy, U. B., & Akinoglu, O. (2012). The effectiveness of elaboration and organizational strategies in science education on students' academic achievement, attitude and concept learning. *Energy Education Science and Technology Part B: Social and Educational Studies (Special Issue)*, 212–218.
- *Dawes, M. E., Horan, J. J., & Hackett, G. (2000). Experimental evaluation of self-efficacy treatment on technical/scientific career outcomes. *British Journal of Guidance & Counseling*, 28, 87–99. <http://dx.doi.org/10.1080/030698800109637>.
- *Demircioğlu, H., Demircioğlu, G., & Calik, M. (2009). Investigating the effectiveness of storyline embedded within a context-based approach: The case for the periodic table. *Chemistry Research and Practice*, 10, 241–249. <http://dx.doi.org/10.1039/b914505>.
- *Duartepe-Paksu, A., & Ubuz, B. (2009). Effects of drama-based geometry instruction on student achievement, attitudes, and thinking levels. *The Journal of Educational Research*, 102, 272–286. <http://dx.doi.org/10.3200/JOER.102.4.272-286>.
- *Eskandar, F. A., Bayrami, M., Vahedi, S., & Ansari, V. A. A. (2013). The effect of instructional analogies in interaction with logical thinking ability on achievement and attitude toward chemistry. *Chemistry Education Research and Practice*, 14, 566–575. <http://dx.doi.org/10.1039/C3RP00036B>.
- Halat, E., Jakubowski, E., & Aydin, N. (2008). Reform-based curriculum and motivation in geometry. *Journal of mathematics, science and technology education*, 4, 285–292.
- *Hong, H. Y., & Lin-Siegler, X. (2012). How learning about scientists' struggles influences students' interest and learning in physics. *Journal of Educational Psychology*, 104, 469. <http://dx.doi.org/10.1037/a0026224>.
- *Hwang, G. J., Wu, P. H., & Ke, H. R. (2011b). An interactive concept map approach to supporting mobile learning activities for natural sciences courses. *Computers & Education*, 57, 2272–2280.
- *Isiksal, M., & Askar, P. (2005). The effect of spreadsheet and dynamic geometry software on the achievement and self-efficacy of 7th-grade students. *Educational Research*, 47, 333–350. <http://dx.doi.org/10.1080/00131880500287815>.

- *Kara, Y., & Yeşilyurt, S. (2008). Comparing the impacts of tutorial and edutainment software programs on students' achievements, misconceptions, and attitudes towards biology. *Journal of Science Education and Technology*, 17, 32–41. <http://dx.doi.org/10.1007/s10956-007-9077-z>.
- *Kazeni, M., & Onwu, G. (2013). Comparing effectiveness of context-based and traditional approaches in teaching Genetics: Student views and achievement. *African Journal of Research in Mathematics, Science and Technology Education*, 17, 50–62. <http://dx.doi.org/10.1080/10288457.2013.826970>.
- Ketbriichi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education*, 55, 427–443.
- *Koksal, E. A., & Berberoglu, G. (2014). The effect of guided-inquiry instruction on 6th grade Turkish students' achievement, science process skills, and attitudes toward science. *International Journal of Science Education*, 36, 66–78. <http://dx.doi.org/10.1080/09500693.2012.721942>.
- *Nwagbo, C. (2006). Effects of two teaching methods on the achievement in and attitude to biology of students of different levels of scientific literacy. *International Journal of Educational Research*, 45, 216–229. <http://dx.doi.org/10.1016/j.ijer.2006.11.004>.
- Ogbuehi, P. I., & Fraser, B. J. (2007). Learning environment, attitudes and conceptual development associated with innovative strategies in middle-school mathematics. *Learning Environments Research*, 10, 101–114. <http://dx.doi.org/10.1007/s10984-007-9026-z>.
- *Özmen, H. (2008). The influence of computer-assisted instruction on students' conceptual understanding of chemical bonding and attitude toward chemistry: A case for Turkey. *Computers & Education*, 51, 423–438. <http://dx.doi.org/10.1016/j.compedu.2007.06.002>.
- *Riconscente, M. M. (2013). Results from a controlled study of the iPad fractions game motion math. *Games and Culture*, 8, 186–214. <http://dx.doi.org/10.1177/1555412013496894>.
- *Ritchie, S. M., Tomas, L., & Tones, M. (2011). Writing stories to enhance scientific literacy. *International Journal of Science Education*, 33, 685–707. <http://dx.doi.org/10.1080/09500691003728039>.
- *Spinner, H., & Fraser, B. J. (2005). Evaluation of an innovative mathematics program in terms of classroom environment, student attitudes, and conceptual development. *International Journal of Science and Mathematics Education*, 3, 267–293. <http://dx.doi.org/10.1007/s10763-004-6531-8>.
- *Stoeger, H., Duan, X., Schirner, S., Greindl, T., & Ziegler, A. (2013). The effectiveness of a one-year online mentoring program for girls in STEM. *Computers & Education*, 69, 408–418. <http://dx.doi.org/10.1016/j.compedu.2013.07.032>.
- *Tarim, K., & Akdeniz, F. (2008). The effects of cooperative learning on Turkish elementary students' mathematics achievement and attitude towards mathematics using TAI and STAD methods. *Educational Studies in Mathematics*, 67, 77–91. <http://dx.doi.org/10.1007/s10649-007-9088-y>.
- *Tsuei, M. (2012). Using synchronous peer tutoring system to promote elementary students' learning in mathematics. *Computers & Education*, 58, 1171–1182. <http://dx.doi.org/10.1016/j.compedu.2011.11.025>.
- *Tuan, H. L., Chin, C. C., Tsai, C. C., & Cheng, S. F. (2005). Investigating the effectiveness of inquiry instruction on the motivation of different learning styles students. *International Journal of Science and Mathematics Education*, 3, 541–566. <http://dx.doi.org/10.1007/s10763-004-6827-8>.
- *Weisgram, E. S., & Bigler, R. S. (2007). Effects of learning about gender discrimination on adolescent girls' attitudes toward and interest in science. *Psychology of Women Quarterly*, 31, 262–269.
- *Welch, A. G. (2010). Using the TOSRA to assess high school students' attitudes toward science after competing in the FIRST Robotics competition: An exploratory study. *Eurasia Journal of Mathematics, Science & Technology Education*, 6, 187–197.
- *Wendell, B. K., & Rogers, C. (2013). Engineering design-based science, science content performance, and science attitudes in elementary school. *Journal of Engineering Education*, 102, 513–540. <http://dx.doi.org/10.1002/jee.20026>.
- *Yalcinalp, S., Geban, Ö., & Özkan, I. (1995). Effectiveness of using computer-assisted supplementary instruction for teaching the mole concept. *Journal of Research in Science Teaching*, 32, 1083–1095. <http://dx.doi.org/10.1002/tea.3660321007>.