Received: 16 April 2020 | Accepted: 27 April 2022

DOI: 10.1111/bjep.12512

ARTICLE

the british psychological society
promoting excellence in psychology

Reducing gender differences in student motivational-affective factors: A meta-analysis of school-based interventions

Kaley Lesperance[1] | **Sarah Hofer**[2] | **Jan Retelsdorf**[3] | **Doris Holzberger**[1]

[1]Centre for International Student Assessment, Professorship for Research on Learning and Instruction, TUM School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

[2]Faculty of Education and Educational Psychology, Ludwig Maximilian University of Munich, Munich, Germany

[3]Educational Psychology, Faculty of Education, University of Hamburg, Hamburg, Germany

Correspondence
Kaley Lesperance, Centre for International Student Assessment, Professorship for Research on Learning and Instruction, TUM School of Social Sciences and Technology, Technical University of Munich, Marsstr. 20-22, 80335 Munich, Germany.
Email: kaley.lesperance@tum.de

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: ZIB2022; The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany, Grant/Award Number: ZIB2022

Abstract

Background: Research shows that gender differences tend to exist in student motivational-affective factors in core subjects such as math, science or reading, where one gender is stereotypically disadvantaged.

Aims: This study aimed to investigate strategies that could reduce these gender differences by conducting a meta-analysis on school-based intervention studies that targeted student motivational-affective factors. We therefore evaluated whether interventions had differential effects for male and female students' motivational-affective factors in a given academic subject. We also evaluated potential moderator variables.

Method: After conducting a systematic database search and screening abstracts for inclusion, we synthesized 71 effect sizes from 20 primary studies. All included studies were conducted in science or mathematics-related subjects, which are stereotypically female-disadvantaged.

Results: While the interventions had significant positive effects for both genders, there was no statistically significant difference between the two genders with regard to the intervention effects on motivational-affective factors. However, the descriptive effect size for female students (g = .49) was far greater than for male students (g = .28). Moderator analyses showed no significant effects for grade

level, intervention duration, or school subject, but there was a significant influence of intervention method used.
Conclusions: This study demonstrated that school-based interventions have positive effects on motivational-affective factors for both genders. It also provides evidence that interventions in subjects where female students are stereotypically disadvantaged may have greater effects for females than for males. Implications and suggestions for future research are discussed.

KEYWORDS
affect, gender differences, interventions, meta analysis, motivation, students

BACKGROUND

Male and female students often display differences in motivational-affective factors within educational contexts, with either one gender or the other being disadvantaged relative to the domain in question (Wigfield et al., 2002). Many constructs fall under the umbrella of motivational-affective factors. Murphy and Alexander (2000), for example, provide a categorization for motivational constructs. They differentiate between goals, interest, motivation (intrinsic and extrinsic), and self-schema (agency, attribution, self-competence, and self-efficacy). In turn, Pintrich (2003) points to expectancy, value, and affective variables as components of motivation. While in Pintrich's (2003) conceptualization, affective variables are regarded as a subcategory of motivation, variables with an emotional aspect such as enjoyment, anxiety, or boredom also have their own distinction in the literature (Pekrun et al., 2011). All of these factors have been shown to be strongly related to career choices, achievement, and performance outcomes in students (Kim & Pekrun, 2014; Möller et al., 2020; Pintrich, 2003; Wigfield & Eccles, 2000). Research has found that many of these factors strongly predict school performance and academic choices, often above and beyond IQ (Goetz & Hall, 2013; Köller et al., 2001; Parker et al., 2014; Steinmayr & Spinath, 2009). For example, in a review by Rosen et al. (2010), which examined 45 studies on motivation, 27 studies on self-efficacy, and 42 studies on academic self-concept, findings indicated that all of these measures were strongly connected to academic achievement in students from kindergarten to 12th grade. Large-scale studies on affective factors have also demonstrated that pleasant emotions such as enjoyment and pride are positively related to academic achievement, whereas unpleasant emotions such as anxiety are most often negatively related (Pekrun et al., 2002, 2017). Given their strong connection to important academic outcomes, gender differences in motivational-affective factors are concerning.

Gender differences in educational contexts

There is a plethora of evidence demonstrating that gender differences between many student motivational-affective factors exist in various subjects. Results from large-scale international studies such as the Program for International Student Assessment (PISA) provide an overview of student outcomes across many countries. In science, results from PISA 2015 revealed that, on average, across all 72 countries assessed, boys were more likely than girls to report higher intrinsic motivation for science, and greater interest and enjoyment in most science-related topics (OECD, 2016). These findings can be seen in smaller studies as well, with boys displaying more positive science attitudes and a higher likelihood of pursuing a science-related career (Jones et al., 2000; Miller et al., 2006; Weinburgh, 1995). A similar trend exists in mathematics, where boys tend to report higher self-efficacy and more positive attitudes, whereas girls display higher levels of math anxiety and are more likely to perceive math as a low-value subject (Else-Quest et al., 2010; OECD, 2013; Pajares, 2005). On the contrary, in the domain of reading, writing, and language arts, this trend is reversed, with males displaying lower reading and writing self-concepts, more negative attitudes towards reading, and low value beliefs for reading as a subject (Durik et al., 2006; Logan & Johnston, 2009; Marinak & Gambrell, 2010; OECD, 2019; Schleicher, 2019). Given these findings, it comes as no surprise that there continues to be large gaps between males and females in career goals and choices. In the areas of science, technology, engineering, and mathematics (STEM), females are extremely underrepresented in higher education and the labour market (Burke & Mattis, 2007; Dasgupta & Stout, 2014), a pattern that is already evident from career expectations and interest during the school years (Master et al., 2017; Sadler et al., 2012; Shapiro & Williams, 2012). On the other hand, careers in areas such as primary education and healthcare professions display much higher numbers of females than males (Hsu et al., 2010; OECD, 2014, 2019).

The role of gender stereotypes for the development of gender differences

Many research studies have investigated why these gaps between male and female students emerge in different subjects. Gender differences in educational contexts seem to arise as an individual grows and interacts with their environment, and gender stereotypes acquired from the social environment, such as from parents (Casad et al., 2015; Tiedemann, 2000), teachers (Muntoni & Retelsdorf, 2018), and peers (Muntoni et al., 2020) seem to play a large role in the emergence of these differences. There are many different assumed mechanisms of how these gender stereotypes are learned and acquired, such as model learning, reinforcement of gender-typical behaviour, different treatment of boys and girls, or direct expression of gender stereotypical expectations (Gunderson et al., 2012; Heyder et al., 2019). These acquired stereotypes can then have an effect on how individuals process and categorize information, as well as on their choices, behaviours, and beliefs (Martin & Halverson, 1981). As certain school subjects such as math, science, reading, and language arts are often stereotyped towards one gender or another (e.g. math is typically a "boy" subject, reading is typically a "girl" subject), these beliefs are also incorporated into traditional gender stereotypes (Leaper, 2015; Plante et al., 2013; Schmenk, 2004). Eccles' expectancy-value theory provides a promising explanation for the mechanisms by which these stereotypes can affect individual choices and behaviours, stating that whether or not an individual undertakes a task depends on their expectations for success and how valuable they perceive the task to be (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). Gender stereotypes can shape not only how valuable a task is to an individual, but also their self-concept, attitude, and perceived competence in that task, which in turn affects their expectancy for success (Eccles, 1994). According to this framework, if a girl perceives math as a male-associated subject, she will not only place low value in it but will also perceive herself as less competent in math, therefore expecting less success, and putting less effort into math or not choosing to study math later in life. Through these mechanisms, gender stereotypes then lead to differences in what young girls and boys are interested in and enjoy, their beliefs about their own capabilities, and the choices they make throughout their academic careers (Eccles et al., 1993; Wigfield & Eccles, 1994).

School-based interventions to reduce gender differences?

Ensuring equal opportunities in education means striving for students to learn and develop according to their full potential. Investigating strategies to reduce gender differences in motivational-affective factors is a step forward in the effort to help all students thrive and succeed. With regard to motivational-affective factors of students in general, regardless of gender, an increasing number of researchers have called for studies to develop methods that could positively reinforce or strengthen these factors and thereby positively influence student achievement outcomes (Gutman & Schoon, 2014; Heckman et al., 2006; Lleras, 2008). There is evidence that motivational-affective factors remain malleable throughout an individual's lifespan, and can therefore be built upon and changed through experience and individual development (Heckman & Kautz, 2013). One promising way of achieving this is through school-based interventions. For the purpose of this study, we define "school-based interventions" as any method used in a school-context, which is different from regular instruction, including not only in-class interventions but also novel teaching methods, summer school programmes, or school-organized workshops. Many studies have empirically tested a variety of these interventions to evaluate the effects on student motivational-affective factors. Previous research syntheses have aggregated the effects of some of these interventions. For example, Durlak et al. (2011) performed a meta-analysis on 213 school-based interventions targeting student factors such as attitudes and emotional skills, and found that overall, students who participated in the interventions significantly improved on measures of these outcomes compared to controls. Additionally, in a review by Gutman and Schoon (2014), results showed that factors such as motivation and self-perceptions of children and adolescents were positively affected in all intervention settings. However, these research syntheses did not evaluate the differential effects of the interventions regarding gender and school subject. It is still unclear whether certain intervention methods have stronger effects on males or females, or whether any of these methods are effective in reducing the differences between male and female student motivational-affective factors in a given academic domain. Additionally, these studies focused on a broader range of student factors (such as social skills, conduct problem, social behaviour, self-control, and creativity), not specifically motivational-affective factors, and did not consider possible moderator variables such as student grade level or intervention duration in their analyses.

The methods used by these intervention studies, as well as the targeted motivational-affective factors and student groups, vary widely. For the purposes of this study, intervention studies can be separated into two types: those that target motivational-affective factors in students overall (i.e. no specific gender is targeted, no gender-specific hypotheses) and those that target motivational-affective factors in one gender specifically, generally the gender that is typically disadvantaged in a given subject (i.e. gender-specific hypotheses). Within these two categories, many different interventional methods are used. Some of these methods can be classified as "psycho-social interventions", while others are more related to classroom processes. Psycho-social interventions are designed to directly target students' subjective psychological processes in an attempt to positively alter them (Walton, 2014). These interventions use strategies such as value affirmations, reframing techniques, and mindset changes. One example of psycho-social interventions without gender-specific hypotheses are utility-value interventions, where students are asked to relate the information they learn in class to their everyday lives in an effort to increase interest for the subject (Hulleman et al., 2010). Psycho-social interventions have also been used for gender-specific interventions by targeting the disadvantaged gender in certain subjects to directly challenge gender stereotypes that students hold. One such strategy is exposing students to role models or mentors who occupy non-traditional gender roles (e.g. a female engineer or a male nurse). Exposure to an individual in a non-traditional gender role can challenge gender stereotypes, lessening the effects of these stereotypes on student self-beliefs (Morgenroth et al., 2015). Another strategy is to target gender-specific student value beliefs. According to Eccles (1994), students will be more likely to engage and put forth effort in a subject if they perceive it to be valuable. While students may find little value in subjects that stereotypically do not align with their gender identity, strategies to make the subject material personally relevant to these students can change how important they view it to be, thereby increasing the likelihood that they engage, take interest, and continue studying that subject (Hulleman et al., 2010).

Outside of psycho-social interventions, other interventional methods focus more on the instructional processes in the classroom, using diverse teaching methods to promote higher levels of motivational factors or more positive affect for students in general. Active learning strategies such as cooperative learning or problem-based learning have been shown to increase student engagement, motivation, and self-efficacy in various settings (Bruder & Prescott, 2013; Laal & Ghodsi, 2012). These strategies attempt to engage students socially and cognitively, encouraging them to be active rather than passive learners, thereby increasing their interest, motivation, and enjoyment

(Hmelo-Silver, 2004; Slavin, 2011). Additionally, instructional methods that integrate digital media, such as interactive online lessons or digital games, are being evaluated more frequently as possible strategies for positively affecting student motivational-affective factors (Erhel & Jamet, 2013; Lieberman et al., 2009; Wang & Reeves, 2007). Using digital media in instruction can also affect student motivational-affective factors by providing a novel environment for learning and possibilities for adapting to individual learner needs and interests (Annetta, 2008; Christensen, 2002; Uzunboylu & Karagozlu, 2015). While these methods mostly target student motivational-affective factors in general, pre-existing differences between male and female students in certain subjects could lead to differential effects of these interventions as well.

In sum, an increasing number of studies have evaluated the effects of school-based interventions on student motivational-affective factors. However, it remains unclear whether these interventions have different effects for female and male students in a given subject, and which are most effective regarding gender-specific deficits. Additionally, these studies vary widely in the interventional methods and motivational-affective factors they address, as well as the age and grade level of the student population. The duration and implementation of these interventions also differ between studies. It is unclear which variables can influence the intervention effects concerning gender differences. Therefore, one goal of the present meta-analysis is additionally to evaluate potential moderator variables of the intervention effects.

Potential moderator variables

Theoretical moderators

School subject

Motivational-affective factors are often closely linked to academic domains (e.g. mathematics, science, etc.) and can therefore vary across these domains (Marsh et al., 2001; Wigfield et al., 2004). Student characteristics such as self-concept, motivation, interest, self-efficacy, enjoyment, and anxiety have all been shown to be tightly connected to the subject they are measured in (Bong & Clark, 1999; Goetz et al., 2006; Green et al., 2007; Wigfield et al., 2004). Therefore, a student might have high self-concept in mathematics, but low self-concept in reading. The fact that these factors are domain specific is also an important characteristic with regard to gender differences between students. Due to the gender stereotypes associated with certain academic subjects, gender differences also vary with regard to the domain in question (for example, boys are usually disadvantaged in reading, while girls are usually disadvantaged in science), and must therefore be discussed in a domain-specific context. Therefore, we included the school subject as a moderator.

Gender-targeted vs. non-targeted

The studies included in this meta-analysis evaluate interventions that target student motivational-affective factors *in general* and therefore have no gender-specific hypotheses. We also include interventions that target the motivational-affective factors of a *specific gender* and therefore hypothesize that the intervention will have differential effects for males and females. A "non-targeted" intervention aims to positively affect the motivational-affective factors of all students, while a gender-targeted intervention aims to positively affect the motivational-affective factors of a specific gender, usually the stereotypically disadvantaged gender in a given subject (e.g. girls in science or boys in reading). Therefore, we included this as a moderator variable.

Grade level

Intervention effects may also vary as a function of student grade level. While motivational-affective factors retain a malleable quality throughout an individual's life, there is evidence that they are more flexible at earlier ages (Gutman & Schoon, 2014). Children's self-perceptions have been shown to decline with age from first grade to 12th grade (Eccles et al., 1993; Jacobs et al., 2002). Additionally, there is evidence that children endorse traditional stereotypes more as they grow older (Rowley et al., 2007) and gender stereotypical self-beliefs in school have been shown to take effect around grade three (Herbert & Stipek, 2005). Therefore, we also explored grade level as a moderator variable.

Intervention duration

Some previous reviews on intervention studies have found that the duration of the intervention may affect how successful the intervention is. Hattie et al. (1996), found a small effect for intervention duration, with shorter interventions (1 or 2 days) having a greater initial impact, but with longer interventions (4–30 days) being more effective overall. Additionally, in their review of the effects of reading and mathematics programmes on student performance, Slavin and Lake (2009) found that interventions with briefer durations reported somewhat larger effect sizes than those with longer durations. Other meta-analyses on intervention studies have also found that the intervention duration could be a moderator of intervention effects (de Boer et al., 2014; Dignath & Büttner, 2008). Given these findings, we included intervention duration as an additional moderator variable.

Robustness moderators

Study quality

In addition to these theoretical moderators, we included study quality as a methodological moderator to control for the effect of study quality on effect sizes. We used the What Works Clearinghouse Standards for Intervention Studies (What Works Clearinghouse, 2020) as a guide when selecting criteria for study quality. To be accepted for inclusion, studies had to compare an experimental group and a control group, and use either random or quasi-experimental assignment. Studies also had to provide pre- and post-test data of participants. Additionally, we coded certain aspects of the instruments used by the studies to measure the outcome variable, namely whether the instrument was established or self-developed by the researchers, and whether the reliability was high, low, or not reported, according to standard rules of thumb for instrument reliability (Tavakol & Dennick, 2011). In addition to these study quality moderators, we also used pre-test data to assess the baseline equivalency for the control and intervention groups.

Type of motivational-affective factor

We also included the type of motivational-affective factor measured as a moderator in order to be able to investigate differences and commonalities between types of motivational-affective factors. For motivational-affective factors, we considered attitudes, beliefs, expectations, motivation, career aspirations, interest, self-concept, self-efficacy, self-confidence, enjoyment, boredom, engagement, anxiety, and satisfaction as possible outcomes.

The present meta-analysis

The main goal of the present research is to conduct a meta-analysis on studies that tested the effects of interventions on student motivational-affective factors and reported gender-specific results. This allowed us to evaluate the following research questions:

1. Do school-based interventions that promote motivational-affective factors in students have differential effects for the stereotypically disadvantaged gender (e.g. males in reading/language arts and females in STEM) and stereotypically non-disadvantaged gender in a given school subject?
2. Are the effects of school-based interventions moderated by:
   a. whether the intervention is gender-targeted or non-targeted?
   b. the grade level of the students?
   c. the intervention duration?

Given the existing evidence on gender differences in student motivational-affective factors, we expect that school-based interventions that target these factors in students will have differential effects for males and females, in particular, more positive effects on the gender typically affected by effects of negative stereotypes in a given school subject (Research Question 1). We also expected that the intervention target (gender targeted vs, non-targeted), grade level and intervention duration would moderate the differential effects of the intervention on students' motivational-affective outcomes (Research Question 2). We also included type of motivational-affective outcome and intervention method as moderators in order to be able to identify specific effects of the various interventions.

In order to assess the above questions, we are interested in both the absolute intervention effects for male and female students (i.e. the intervention effects for girls or boys respectively) as well as the intervention effects on the difference between girls and boys (i.e. if pre-existing gender differences are significantly reduced by the intervention).

METHOD

Literature search and study selection

We first conducted a literature search in the databases of PsycINFO and Educational Resources Information Center (ERIC). A flowchart of the study selection process can be seen in Figure 1. The keywords used pertained to (1) the population of interest (students, school, etc.), (2) the topic of gender differences, (3) school-based interventions, and (4) motivational-affective student factors of interest. Keywords varied slightly between the two databases based on the subject heading classification system of each database. The full syntax is included in Appendix A. We restricted the search to studies written in English. In order to include grey literature, we did not restrict the search by publication type, therefore including grey literature such as dissertations, conference proceedings, and other literature formats. We did not restrict the search results by year published in order to search all studies on this topic as thoroughly as possible. This first search was conducted on 13 May 2019. The databases searches resulted in 5,650 references (after removing duplicates), with 4,480 results from ERIC and 1,170 results from PsycINFO. The searches from ERIC and PsycINFO were updated on 26 August 2020, resulting in an additional 184 new results from ERIC and 38 new results from PsycINFO. An additional search in the Web of Science database was also conducted on this date, which returned 2,607 results. Additionally, the reference lists of all included studies, as well as existing meta-analyses or reviews on similar topics, was manually screened for possibly relevant articles, which resulted in an additional 28 studies. In order to locate unpublished studies, the authors of all included studies were contacted via email to inquire about any additional unpublished works that might be relevant. This resulted in one additional study. We also sent a call for papers and/or data to the mailing lists and newsletters of the European Association for Research on Learning and Instruction (EARLI), the American Educational Research Association (AERA), the European Educational Research Association (EERA), and the Gender and STEM research network. This call for papers detailed the topic of interest of the meta-analysis, as well as the study inclusion criteria, and invited researchers to send any relevant published or unpublished work. However, we received no responses from this call.

The titles and abstracts of all studies were screened using the following inclusion criteria:

1. **Study Design**. Studies were only included if they compared an experimental group to control groups, with either random or non-random (i.e. clustered) assignment to groups. Studies also were required to include pre-test data.
2. **School Level**. Only studies conducted with students at the primary or secondary school level were included. Studies in higher education or pre-school levels were excluded.

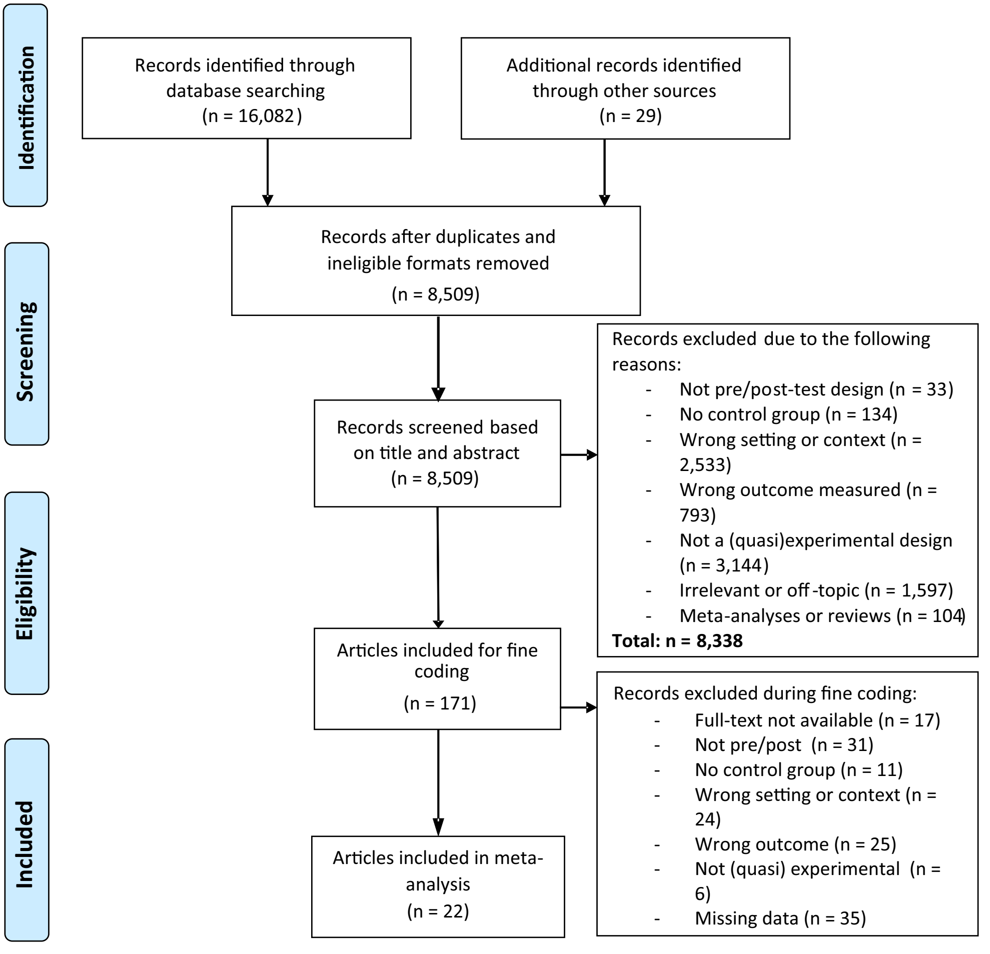

FIGURE 1 Flow chart of the study selection process. Study selection was done following the guidelines of The PRISMA Group (Moher et al., 2009)

3. **School Subject**. Only studies conducted in the core subjects of mathematics, science, reading/(native)language arts or STEM were included. Studies conducted in alternative subjects, such as physical education, arts, or foreign language, were not considered.
4. **Sample Composition**. Only studies that had a sample that represented the average student population of the class were included. Studies with samples consisting exclusively of a specific ethnic or religious group were excluded, if the sample was purposefully selected out of the general school population, for example, if only the African American males in the school were included in the study, or only Latin-American girls were selected to be in the sample (this was not the case if the study takes place in a country where the sample is naturally made up of a specific ethnic group, for example, Saudi Arabia or Mexico). Along the same lines, studies with samples consisting exclusively of gifted or special education students were also excluded. These subgroups are often described as having specific characteristics (based on their background and prior experiences) that are in the focus of corresponding interventions. Accordingly, the results of these intervention studies cannot be transferred to a general student sample.
5. **Intervention Study**. Only studies that evaluated a school-based intervention were included. Interventions conducted solely in the home environment (e.g. with parents or siblings) or outside of a school context (e.g. in church, in sports clubs, etc.) were not included.

6. **Motivational-Affective Factors as Outcome**. Studies were included only if they evaluated the effects of the intervention on one or more motivational-affective student factors.

Using the information from the titles and abstracts of each reference, articles were categorized as “Included” or “Excluded”, with a reason given for each exclusion with regard to the criteria. A portion (275) of the articles were double-screened by two separate, trained research assistants. The double coding showed that the inclusion and exclusion process was reliable, with 91% agreement between the coders' overlapping articles. Disagreements were discussed to arrive at a final decision.

Coding of data

In order to code all relevant study variables, full texts were acquired for the studies that were included. The studies were coded using a coding scheme developed by the researchers. The coding scheme was created according to the general publication features (author, date, etc.), as well as methodological variables such as the intervention duration, implementation (instructor of intervention, setting of intervention, etc.), measurement instruments, content-related variables such as the theoretical background, sample characteristics, and school subject, and quantitative data necessary for effect size calculations such as the means, standard deviations, and sample sizes. In order to assess the gender-specific intervention effects, we coded the pre-test and post-test scores for the experimental and control groups for the entire sample, for the female participants only, and for the male participants only. An overview of all variables coded and their purpose can be seen in Table 1. In order to ensure coding reliability, a portion of the included studies were fully double coded by two trained researchers to assess inter-rater agreement. As an indicator for inter-rater reliability, Cohen's Kappa (Cohen, 1968) was calculated for each variable. Values for Cohen's Kappa ranged from $\kappa = .51$ (intervention method), which is considered a substantial agreement, to $\kappa = 1.00$ (grade level), which represents perfect agreement (Landis & Koch, 1977). The agreement for variables without any margin for interpretation, such as means, standard deviations, and sample sizes for effect size calculation, was almost perfect. Any disagreements were discussed between the two coders. During the fine coding process, it was necessary to exclude some studies which originally seemed eligible, as upon further inspection, they were either not a fit for the meta-analysis or did not report the necessary information (the most common case being that the authors had not evaluated gender differences). In the case of missing information, authors of the respective articles were contacted, when possible, to request missing data. In total, we fully coded 22 eligible studies.

Moderator variables

We originally planned to include all moderator variable in one meta-regression model. However, due to the smaller number of studies we included than anticipated, we instead decided to conduct a separate

TABLE 1 Overview of coded variables

Study identification	Descriptive variables	Study quality	Effect size and analyses	Moderator variables
Author(s) Title Year of publication Document type	Intervention method Country	Study design Source of instrument Instrument reliability	Sample sizes Single sex or both sexes Means (pre/post) Standard deviations (pre/post)	Grade level Intervention duration Motivational-affective factor Target of intervention

meta-regression for each moderator in order to assess the effects. For the categorical moderators, we added dummy-coded predictors for each of the different levels.

School subject

We originally coded the school subject for each study as stated in the study (e.g. Biology was coded as "Biology" and not "Science"). However, after completing the coding process, we categorized the different subjects as Science, Mathematics, Reading, and Informatics/Technology (therefore, "Biology" would now be coded as "Science"). Upon completion of the fine coding, we were only left with one study conducted in a "male-disadvantaged" subject (Kerneža & Košir, 2016, conducted in reading literacy). All other studies were conducted in the subjects of science or math, with two exceptions conducted in technology/computer science. We therefore excluded the study by Kerneža and Košir (2016), as one study in a male-disadvantaged subject would not allow us to accurately assess the differential effects. We therefore ended up with only studies that were conducted in the typically female-disadvantaged STEM subjects of science, math, and technology.

Gender targeted vs. non-targeted

An intervention was coded as gender targeted when the authors stated that they expected the intervention to have different effects for male and female students, either in their research questions or hypothesis. An intervention was coded as not gender-targeted when no gender-specific effects were considered in the research question or hypotheses.

Grade level

We originally coded grade level according to the specific grade reported in the study (e.g. Grade 3, Grade 10). If this was not reported, we coded the school level reported in the study (e.g. primary school, high school). Once the coding was complete, we categorized each grade level into Primary (Grades 1–5), Lower Secondary (Grades 6–8) or Upper Secondary (Grades 9–12).

Intervention duration

We used the number of weeks to represent the intervention duration. The shortest duration was 1 week. We centred this variable by subtracting 1 from each value, in order to have a meaningful intercept in the meta-regression model (so the intercept would represent a one-week intervention).

Intervention method

In order to assess which intervention methods were most affective, we classified the interventions into two categories according to our previous description of intervention types. The first category was psycho-social interventions. These interventions directly targeted students' motivational-affective processes by attempting to change or restructure their subjective beliefs and perceptions. This category included interventional methods such as utility-value interventions, role models, and challenging stereotypes. The second category was instructional interventions. These interventions included strategies that focused on the instructional and pedagogical processes in the classroom, with the aim of using diverse teaching methods to change students' motivational-affective factors. This category included interventional methods such a problem-based learning, cooperative learning, or novel curriculum designs. We categorized interventions according to the theoretical frameworks and designs used in the primary studies.

Study quality

We originally planned to use study design (experimental vs. quasi-experimental), statistical reliability of the instrument used to measure outcomes (high vs. low vs. not reported), and the source of the instrument (whether it was an already existing instrument or developed by the researchers of the corresponding study themselves) as variables to evaluate study quality. However, upon completion of the coding, it became evident that all included studies were quasi-experimental, and we therefore could no longer use this variable as a moderator. We coded statistical reliability of the instrument as high, low, or not reported, and the instrument source was coded as an already existing instrument or one that the researchers created themselves. After completing the coding process, only one study was included which did not report the instrument reliability, while all other studies reported high reliability. We also found, upon completing the coding process, that all but one study included used an already existing instrument. We therefore did not include these variables as moderators. We calculated baseline equivalency for the included samples as the standardized mean difference between the control and experimental groups at pre-test, as per the What Works Clearinghouse standards (What Works Clearinghouse, 2020).

Type of motivational-affective factor

While we originally planned to assess the effects of the interventions on each separate motivational-affective factor individually, the final data did not allow enough power for this type of analysis. We therefore classified the motivational-affective factors into four different categories, loosely based on the framework of Murphy and Alexander (2000). The four categories were (1) motivational factors (motivation, interest, value, and engagement), (2) self-schemas (self-efficacy, self-concept, self-confidence, and stereotypes), (3) attitudes, and (4) affective factors (enjoyment, anxiety, and satisfaction). We reversed coded the data for outcomes where a decrease in the mean score represents a positive outcome (e.g. anxiety or stereotypes), so that all effect sizes represented the same directional relationship (i.e. a positive effect size indicates desirable change in motivational-affect factor).

Calculation of effects and general analytic strategies

All analyses were conducted in R Version 4.0.5. (R Core Team, 2019), using the *metafor* package (Viechtbauer, 2010) as well as the *robumeta* package (Fisher & Tipton, 2015). All included studies were quasi-experimental designs with pre–post data from treatment and control groups. Therefore, we calculated the effect sizes for males and females as the difference (g) between the standardized mean change score for the treatment and control groups, including a small-sample bias adjustment. We calculated this using the means, standard deviations, and sample sizes of the treatment and control groups for both males and females in each study. According to Morris (2008), the effect size for pre–post-control design studies is defined as the mean difference between post-test and pre-test scores, divided by the common standard deviation. Following the recommendations made by Morris (2008), we used the pooled pre-test standard deviation to calculate standardized mean change score for pre–post-control studies, as this provides an unbiased estimate of the population effect size. This effect size was calculated as follows:

$$g = c_p \left[\frac{\left(M_{post,T} - M_{pre,T}\right) - \left(M_{post,C} - M_{pre,C}\right)}{SD_{pre}} \right]$$

where the pooled pre-test standard deviation is defined as

$$SD_{pre} = \sqrt{\frac{(n_T - 1)SD^2_{pre,T} + (n_C - 1)SD^2_{pre,C}}{n_T + n_C - 2}}$$

and the bias correction is defined as

$$c_p = 1 - \frac{3}{4(n_T + n_C - 2) - 1}$$

We used random-effects statistical models for this meta-analysis, as this allows the findings to be generalized beyond the included studies (Hedges & Vevea, 1998). We coded at least two effect sizes per study (male vs. female), and if the study measured multiple motivational-affective factors, these were also coded as separate effect sizes. Separate effect sizes that come from the same study are not independent of each other, and therefore violate the independency assumption of classical random-effects meta-analyses (Hedges et al., 2010). We remedied this by using robust variance estimation (RVE) to estimate our model. RVE allows for the inclusion of statistically dependent effect sizes by adjusting the standard errors to account for dependency (Tanner-Smith et al., 2016). We used correlated effect model weights to model the unknown covariance structure (Fisher & Tipton, 2015).

Heterogeneity across the included studies was assessed by calculating the 95% prediction interval, which describes the expected range of true effects by predicting where the true effects are to be expected for 95% of similar future studies (Borenstein et al., 2017). Therefore, this interval can be used to evaluate the variability of intervention effects over different settings (IntHout et al., 2016). To evaluate the first research question, whether the interventions demonstrated a differential effect for male and female students on motivational-affective outcomes in a given school subject, we estimated a simple random effects meta-regression model using RVE, with gender included as a predictor. Due to the lack of studies conducted in a male-disadvantaged subject, we could not evaluate the differential effects of a male-disadvantaged versus female-disadvantaged subject. However, we still included the main subjects of science and mathematics as moderators.

Fourteen of the included studies used a cluster-randomized design, where treatment and control conditions were assigned at the classroom level (i.e. using one pre-existing classroom as the treatment group and another as a control group), but outcomes were reported on the student level. This clustering effect can lead to additional variance, meaning that it was necessary to adjust the variances estimates for studies using this design (Hedges, 2007). As none of the studies reported the intraclass correlation (ICC) necessary for the variance adjustment, we used a conservative estimate of ICC = .20, as recommended by Hedges and Hedberg (2007).

With regard to the second research question, we used separate meta-regression models to investigate the effects of each individual moderator. We originally planned to examine publication bias via selection models. Selection models aim to directly model the selective publication process by considering the probability that certain studies are included in a meta-analysis based on specific characteristics and using weight functions to adjust the overall effect size estimate (Vevea & Hedges, 1995). However, according to McShane et al. (2016), realistic selection models cannot be properly estimated without a large amount of data, and without sufficient data, selection models cannot be relied on to provide accurate estimates. Due to the relatively small final sample size of our meta-analysis, we therefore did not evaluate a selection model. We evaluated publication bias visually by inspecting the funnel plot, as well as statistically using Egger's regression test (Egger et al., 1997) and Kendall's rank correlation test (Begg & Mazumdar, 1994).

In a meta-analysis, outliers and other exceptional cases can affect the interpretability and robustness of results (Viechtbauer & Cheung, 2010). Therefore, we conducted a variety of diagnostic tests on the overall model to examine this. We calculated the: (1) externally standardized residuals, (2) DFFITS values, (3) Cook's distances, (4) covariance ratios, (5) leave-one-out estimates of the amount of heterogeneity, (6) leave-one-out values of the test statistics for heterogeneity, (7) hat values, and (8) weights.

We used the *metafor* package to calculate effect sizes, and the *robumeta* package to estimate our overall model as well as meta-regression models for our moderators, using robust variance estimation to correct for dependent effect sizes. Additional analyses for publication bias, heterogeneity, and outlier detection were all conducted in *metafor*, using a model estimated without correction for dependent effect sizes, as these additional analyses are not possible in *robumeta*. A registered protocol of this study, along with a template of the coding scheme and the R scripts can be found on the Open Science Framework website via the following link: https://osf.io/zb8sc/?view_only=7392757de06e45ffbd3c04c87d569b9c.

RESULTS

Descriptive characteristics

The inclusion criteria were met by 21 independent studies, with 79 relevant effect sizes obtained from a total of 3,458 participants overall. All studies were published or conducted between 2000 and 2019, with the exception of 1 article published in 1981. The sample sizes ranged from 11 to 732 participants. There was a large variation in the intervention techniques. Some focused directly on student motivational-affective factors and attempted to alter these by fostering skills such as goal setting, connecting subject matter to relevant aspects of student lives, or challenging stereotypes through examples and role models. Other interventions incorporated different instructional techniques such as problem-based learning, cooperative learning, or digital games in efforts to increase student motivation, attitudes, and engagement. A descriptive overview of the relevant characteristics from each study can be seen in Table 2.

Outlier analysis

The outlier analysis identified the effect sizes from the Akcay et al. (2010) study as outliers. The effect sizes ranged from $g = 2.30$ to $g = 5.14$, which were much larger than any of the other effect sizes from the other studies. As the presence of outliers can affect the validity and robustness of meta-analytic results, we removed this study from all further analyses.

Overall effect and differences in gender and subject (Research question 1)

After removal of outliers, the final number of studies included in the analyses was 20, with a total of 71 effect sizes. The first research question focused on whether interventions that targeted student motivational-affective factors had differential effects for the stereotypically disadvantaged and non-disadvantaged gender in a certain subject. We first ran a model with just gender as a predictor to examine the overall role of gender on the intervention effects (we ran model with female as the reference category and one model with males as the reference category). The results of this model can be seen in Tables 3 and 4. We found a significant positive effect of the interventions on both male and female student motivational-affective factors; however, there was no statistically significant difference between males and females. The overall descriptive effect for female students, at $g = .49$, was almost double that for male students, which was $g = .28$. The measures of heterogeneity indicated substantial heterogeneity between effects sizes, with $I^2 = 80.33$ ($\tau^2 = .14$), indicating that a large percentage of variance is due to heterogeneity between studies. The 95% prediction interval was .71 to 3.24 for females, and .59 to 2.68 for males, also indicating high heterogeneity. A forest plot for the female and male effect sizes from the various studies can be seen in Figures 2 and 3 respectively.

Due to the lack of studies conducted in male-disadvantaged subjects, we were not able to evaluate if there were differential intervention effects for the stereotypically disadvantaged gender in a given school subject. However, we still included school subject as a moderator variable (with the two categories of either science or mathematics) and gender, as well as the interaction between subject and gender, to evaluate whether there were any significant differences of the intervention effects for the interplay between these two variables. Results of these analyses are displayed in Table 5. A random-effects meta-regression model using robust variance estimation did not reveal any significant difference of the interventions effects moderated by the school subject or its interaction with gender. Descriptively, the estimated effect sizes were highest for females in science $g = .46$, whereas for males in science, the effect size was $g = .30$. For females in math, the estimated overall effect size was $g = .40$, and for males in math, $g = .26$. We also conducted post-hoc subgroup analyses to evaluate the absolute effects on the

TABLE 2 Descriptive overview of included studies

Study	Sample Size (N)	Grade Level	Country	Subject	Intervention Method	Gender Targeted	Outcome Measured	Summary of Results
Abed (2016)	87	Lower secondary	Jordan	Science	Drama in science instruction using physical simulations activities (INST)	Yes (only male students served as sample)	Attitudes towards science	The use of drama-based science teaching brought about positive changes in student's attitudes towards science learning and their understanding of scientific concepts compared to traditional science teaching
Al-Balushi and Al-Aamri (2014)	62	Upper secondary	Oman	Science	Project-based learning (INST)	Yes (only female students served as sample)	Attitudes towards science	Students who participated in the project-based learning group developed more positive attitudes towards science than the control group
Alghamdi (2017)	50	Upper secondary	Saudi Arabia	Science	Cooperative learning model by using the Jigsaw instructional strategy (INST)	Yes (only male students served as sample)	Attitudes towards science	Although the use of Jigsaw as a cooperative learning strategy improved the science achievement of students in the experimental group, the strategy did not change the experimental group student's attitudes in a positive direction
Argaw et al. (2016)	81	Upper secondary	Ethiopia	Physics	Problem-based learning in physics (INST)	No	Motivation (to learn physics)	Students taught with a PBL approach showed better outcomes post-test versus control; however, motivation for both groups remains stagnant, no significant differences of intervention in problem solving and in motivation on male and female students

TABLE 2 (Continued)

Study	Sample Size (*N*)	Grade Level	Country	Subject	Intervention Method	Gender Targeted	Outcome Measured	Summary of Results
Batton (2010)	64	Primary	USA	Math	Cooperative learning based on Vygotsky's social learning and Piaget's concept of knowledge (INST)	No	Mathematics anxiety	Decrease in math anxiety is larger for students in cooperative group in comparison to students in non-cooperative group, decrease in math anxiety for female students in cooperative group only, no difference for male students in decrease of math anxiety in cooperative and non-cooperative group
Cady and Terrell (2008)	26	Primary	USA	Computer science	Positive relationship between frequent technology use and self-efficacy and technology acceptance, daily infusion of technology use into science lessons (PSI)	Yes (only female students served as sample)	Attitudes towards computers (importance, enjoyment, self-efficacy)	Students in intervention group develop higher importance and higher self-efficacy of technology use versus control, no significant differences in computer enjoyment between groups
Chen et al. (2019)	68	Primary	Taiwan	Science	Modified Argument-driven Inquiry (MADI) teaching approach (INST)	Yes	Engagement in learning science	The engagement in learning science of the experimental group was statistically significantly higher than those in the control group. However, no statistically significant changes between boys and girls of the experimental group were found

TABLE 2 (Continued)

Study	Sample Size (N)	Grade Level	Country	Subject	Intervention Method	Gender Targeted	Outcome Measured	Summary of Results
Chiu (2011)	247	Lower secondary	Taiwan	Science	Social cognitive approach with lessons focused on women in sciences and men in humanities (PSI)	No	Attitudes towards learning science (interest, confidence, value)	Initial gender gap favouring boys regarding interest and confidence decreased in experimental group only, initial gender gap favouring boys regarding value decreased and changed from favouring boys to favouring girls in experimental group
Fabian et al. (2018)	52	Primary	Scotland	Math	Mobile learning (based on constructivist learning, collaborative learning, situated learning and others) (INST)	No	Attitudes towards mathematics (enjoyment, self-confidence, value of math)	Students in the experimental group had significantly higher gains in performance in comparison to the control group. However, no significant difference was found in student's attitudes towards mathematics between the groups. No significant gender differences were found either
Falco et al. (2008)	228	Lower secondary	USA	Math	Improving students' self-efficacy beliefs towards learning mathematics through fostering the skills time management, goal setting, study habits, and help-seeking. (PSI)	Yes	Attitudes towards mathematics (self-confidence, value, enjoyment, motivation)	Students in experimental group develop more positive attitudes towards mathematics compared to control group, gains for female students compared to male students were significantly higher

TABLE 2 (Continued)

Study	Sample Size (N)	Grade Level	Country	Subject	Intervention Method	Gender Targeted	Outcome Measured	Summary of Results
Falco and Summers (2019)	88	Upper secondary	USA	STEM	Career group intervention based on self-efficacy theory (PSI)	Yes (only female students served as sample)	STEM self-efficacy and career decision-making self-efficacy	Significant differences were found between the experimental and the control group at post-test and follow-up test for both variables, which means the intervention had a positive impact on participant's career decision self-efficacy and STEM self-efficacy
Fennema et al. (1981)	Not reported	Upper secondary	USA	Math	Videotapes aimed to increase knowledge about sex-related differences in mathematics and improvement of attitudes (PSI)	Yes	Mathematics value, anxiety and stereotypes	Students in the experimental group reported that they would increase the amount of mathematics courses they planned to take. The enrolment data also confirmed that the mathematics enrolment of students in the experimental group increased compared to the control group, indicating that the intervention was effective in increasing women's participation in mathematics
Mavridis et al. (2017)	79	Lower secondary	Greece	Math	Online flexible educational game targeting student attitudes as a supplementary teaching method during mathematics lessons (PSI)	No	Attitudes towards mathematics	Students taught with an educational game approach develop more positive attitudes towards mathematics and higher achievement scores versus control, educational game approach works equally well for male and female student's attitudes
Orabuchi (2013)	155	Lower secondary	USA	Math	Online visual and interactive technological tool based on self-efficacy theory (PSI)	No	Mathematics anxiety, attitudes towards mathematics	No significant differences between groups regarding mathematics performance, mathematics anxiety and attitudes, no gender differences

TABLE 2 (Continued)

Study	Sample Size (*N*)	Grade Level	Country	Subject	Intervention Method	Gender Targeted	Outcome Measured	Summary of Results
Shin et al. (2019)	416	Primary and lower secondary	Korea	Science	Expectancy-value-theory, utility value intervention to positively influence STEM motivation (PSI)	No	Attitudes towards science	Students taught with a utility value intervention develop more positive attitudes towards mathematics versus control, science utility value intervention works equally well for male and female students, no significant differences of intervention on male and female students
Soyibo and Hudson (2000)	77	Upper secondary	Jamaica	Biology	Computer-assisted instruction (CAI) in addition to lecture and discussion (INST)	Yes (only female students served as sample)	Attitudes towards biology and the computer/CAI	The experimental group had significantly better post-test attitudes towards biology and computers/CAI than the control group. The experimental group also significantly outscored the control group in their understanding of reproduction in plants and animals
Starkey (2013)	168	Lower secondary	Kenya	Math	Digital game as a supplementary teaching method during mathematics lessons (PSI)	Yes	Attitudes towards mathematics and situational motivation	Significant increase in mathematics achievement, motivation and attitudes towards mathematics for students who played the digital game versus the control group, digital game approach works significantly better for male students regarding their motivation increase

TABLE 2 (Continued)

Study	Sample Size (*N*)	Grade Level	Country	Subject	Intervention Method	Gender Targeted	Outcome Measured	Summary of Results
Sung et al. (2015)	111	Primary	Taiwan	Math	Technology-based 3D virtual manipulatives and step-by-step progressive teaching for teaching the surface area of composite solids (INST)	No	Attitudes towards mathematics	There was no significant difference in student's attitudes towards mathematics between experimental and control groups, nor between male and female students. The experimental group exhibited better performance compared to the control group
Van der Meij et al. (2015)	61	Lower secondary	Netherlands	Science	Animated pedagogical agents in virtual environments (INST)	Yes	Task relevance (value) and self-efficacy	Students taught with APA develop higher self-efficacy beliefs and knowledge versus control group, significant increase of self-efficacy beliefs for female students in experimental conditions and decrease of self-efficacy beliefs for female students in control conditions, opposite pattern for male students
Zhao et al. (2018)	77	Lower secondary	China	Math	Systematic intervention on collective representations, situational cues and personal characteristics based on the Identity Threat Model (PSI)	Yes (only female students served as sample)	Math-gender stereotypes	The level of math-gender stereotypes in the experimental group was significantly lower than that of the control group after the intervention and at a follow-up test, which indicates that the intervention was effective in reducing math-gender stereotypes among adolescent girls

Note: In the "Intervention Method" column, we indicated whether we classified the intervention as instructional (INST) or psycho-social (PSI).

TABLE 3 Model with gender as predictor

Predictor	Estimate	*SE*	95% CI Lower	95% CI Upper	*df*	*p*-value
Intercept	.487	.103	.266	.708	14.4	.0003***
Gender (Male)	−.208	.150	−.529	.108	17.6	.1841

Note: Female is the reference category (intercept). ***$p < .01$.

TABLE 4 Model with gender as predictor

Predictor	Estimate	*SE*	95% CI Lower	95% CI Upper	*df*	*p*-value
Intercept	.279	.126	.005	.554	12.0	.0468**
Gender (Female)	.208	.150	−.108	.524	17.6	.1841

Note: Male is the reference category (intercept). **$p < .05$.

intervention for with females in science, females in math, males in science, and males in math, in order to see the effectiveness of the interventions for each group separately. These results can be seen in Tables A1–A4 in Appendix B.

Differences in effect sizes depending on moderator variables (Research question 2)

In order to investigate our second research question, we conducted various moderator analyses to evaluate whether any potential moderating variables were responsible for the variance among studies. Results for all moderator analyses can be seen in Table 6. For the categorical moderators (i.e. grade level, intervention target, etc.), we chose the categorical level that displayed the descriptively strongest association with the outcome variable as the reference category.

There were no significant differences between the grade levels with regard to the intervention effects. The results descriptively showed the biggest effect size estimate for primary school levels ($g = .48$), followed closely by lower secondary school levels ($g = .43$), and then upper secondary school levels ($g = .32$). We also evaluated the effect of the intervention method (psycho-social vs. instructional) and its interaction with gender on the intervention effects. Results displayed largest effect size estimates for females when psycho-social interventions were used, $g = .53$, whereas the effect size for males when psycho-social interventions were used was significantly lower ($g = .19$). Effects from interventions using instructional interventions, with $g = .42$ for females, and $g = .41$ for males, did not differ from the reference group (effects of psycho-social interventions for females).

For the moderator of "intervention target", we also examined the effects of this variable in combination with gender. In reference to the intercepts, which represented gender-targeted interventions and females, the only significant difference found was for non-gender-targeted interventions and females. The effect sizes for gender-targeted interventions for females was $g = .63$ ($p = .002$) while the effect size for females who received a non-gender-targeted intervention was at $g = .20$ ($p = .015$). The effect size for males who received a gender-targeted intervention was $g = .32$, while the effect size for males who received a non-gender-targeted intervention was $g = .24$. Neither of these conditions was significantly different from the reference category (gender targeted and female).

The duration of the intervention in weeks had no significant moderating effect ($g = .01$). Upon completion of this moderator analysis, we ran an additional post-hoc moderator analysis with duration as a categorical variable. We used the categories of less than or more than 4 weeks, based on a

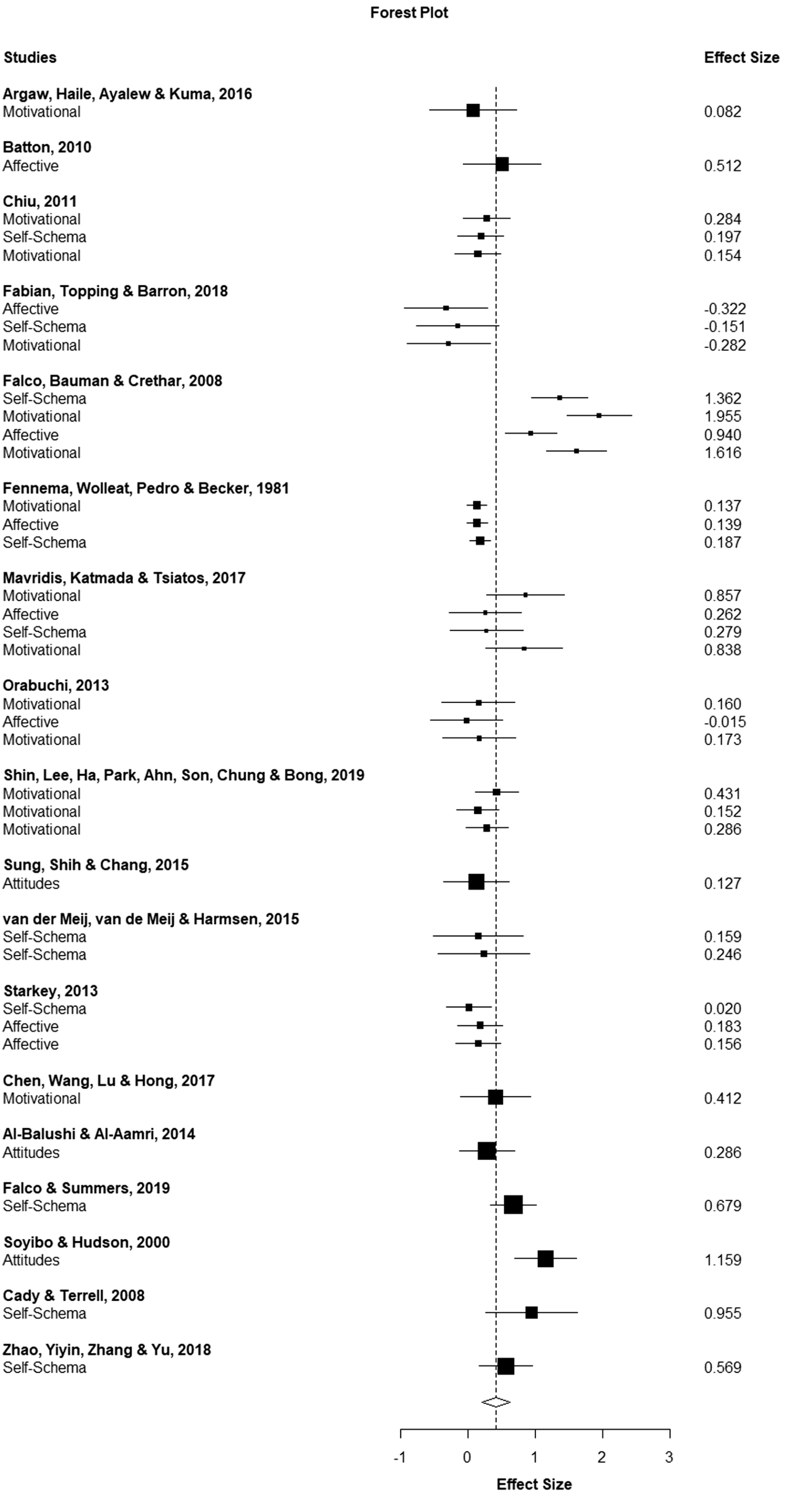

FIGURE 2 Forest plot of female effect sizes. Effect sizes adjusted for dependency using RVE

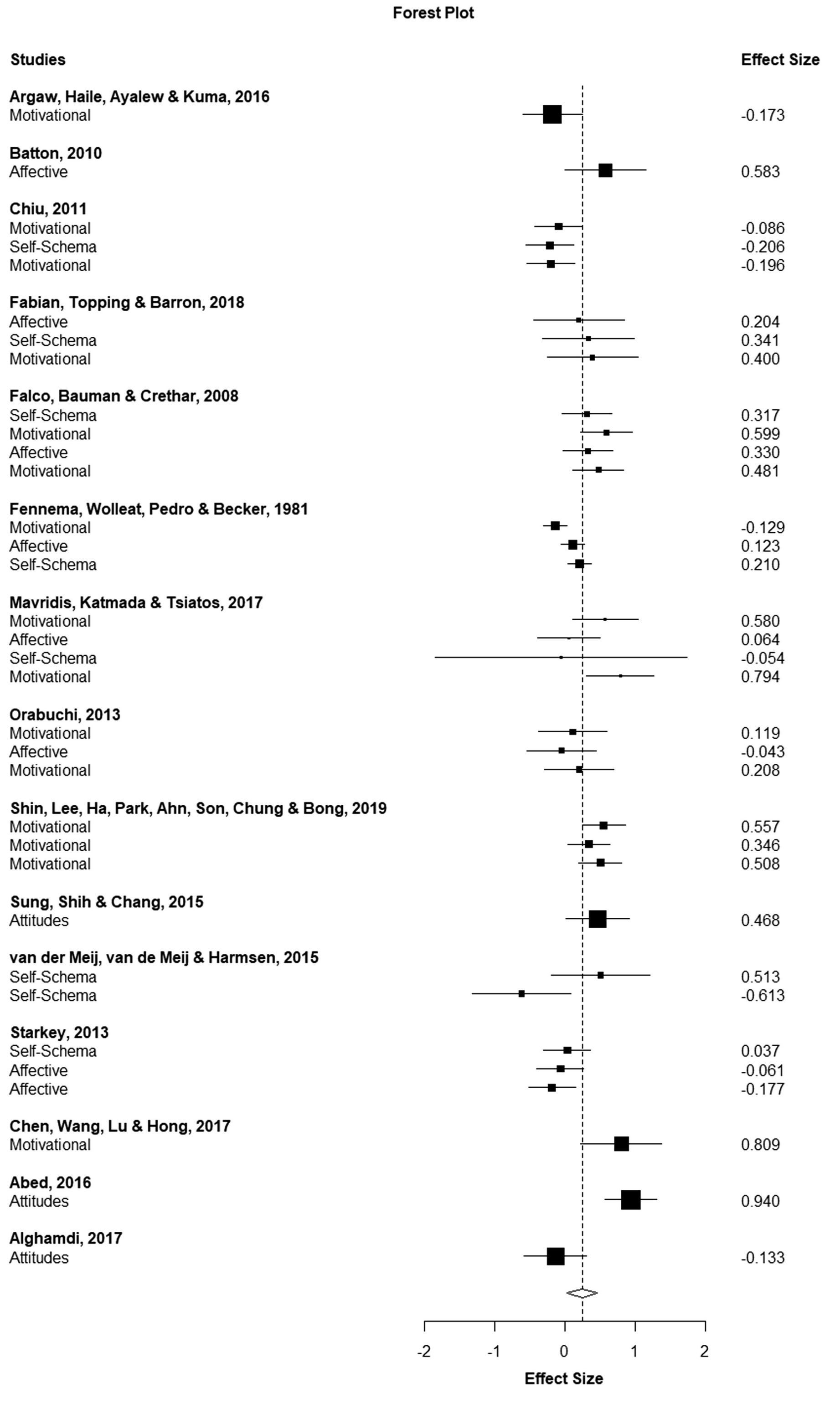

FIGURE 3 Forest plot of male effect sizes. Effect sizes adjusted for dependency using RVE

TABLE 5 Meta-regression with interaction of subject and gender

Moderator Level	*j*	*k*	*df*	Estimate	*SE*	*p*-value	Reference
Subject							
Female & Science	7	12	5.07	.46	.19	.064	REF
Female & Math	9	23	11.51	−.06	.25	.817	
Male & Science	7	12	7.66	−.16	.29	.598	
Male & Math[a]	8	22	15.30	.02	.32	.954	

Note: *j* represents the number of studies, and *k* represents the number of effect sizes. REF indicates which level of the variable was used as the reference category (intercept).

[a]Indicates that this combination was an interaction term in the meta-regression.

meta-analysis by Hattie et al. (1996), which found that interventions under 30 days had a positive correlation with effect sizes. However, this additional analysis also resulted in non-significant results. There were also no significant differences with regard to type of motivational affective outcome measured. Descriptively, the biggest effect sizes could be seen for interventions where attitude was the outcome measured (g = .52), followed by self-schema outcomes (g = .47), motivational outcomes (g = .34), and affective outcomes (g = .24).

Publication bias & robustness checks

Egger's regression test was not significant (p = .960), indicating that there was no evidence for risk of publication bias. The rank correlation test for funnel plot asymmetry was also insignificant (p = .198), also indicating no evidence for publication bias. The funnel plot can be seen in Figure 4.

With regard to study quality, we found that for 18 of the 71 effect sizes, the baseline equivalency standard recommended by the What Works Clearinghouse (an effect size of more than .25) was not met. A large percentage of these problematic baseline equivalencies were within studies, for only one gender but not the other. Therefore, due to our goal to compare the effects of the interventions by gender, we did not exclude these effect sizes from our analyses. However, we conducted a sensitivity analysis to evaluate the effect of baseline equivalency, which can be seen in Table A5 in Appendix C.

DISCUSSION

In the present paper, we aimed to investigate whether interventions that targeted motivational-affective factors in students had differential effects for the stereotypically disadvantaged and stereotypically non-disadvantaged gender in a given school subject. While other meta-analyses and reviews have evaluated the effects of school-based interventions on these factors in students (Durlak et al., 2011; Gutman & Schoon, 2014), the present study takes a closer look at gender-specific effects of these interventions. We also examined additional variables that might moderate the effects of these interventions.

Gender-specific intervention effects

On a general level, the results of this meta-analysis demonstrate that interventions have the potential to promote motivational-affective factors for both male and female students. This is in line with prior research, which shows that school-based interventions can effectively promote or foster these factors in students. Large-scale meta-analyses have demonstrated the positive effect of interventions on student's attitudes, emotional skills, and motivation (Durlak et al., 2011; Lazowski & Hulleman, 2016; Taylor et al., 2017). The findings that these interventions have, on average, a positive effect for both male and

TABLE 6 Overview and results of moderator analyses

Moderator level	*j*	*k*	%	*df*	Estimate	*SE*	*p*-value	Reference category
Grade level								
Primary	5	13	18%	3.96	.48	.14	.029**	REF
Lower secondary	8	42	59%	7.92	−.05	.20	.812	
Upper secondary	7	16	23%	8.17	−.16	.22	.499	
Intervention method								
Psycho-social & female	10	26	37%	7.83	.53	.12	.002***	REF
Instructional & female	8	11	15%	12.02	−.11	.23	.644	
Psycho-social & male	8	23	33%	7.70	−.34	.11	.010**	
Instructional & male[a]	8	11	15%	16.31	.33	.30	.281	
Intervention target								
Gender targeted & female	10	18	25%	7.87	.63	.13	.002***	REF
Non gender targeted & female	8	19	27%	12.37	−.43	.15	.015**	
Gender targeted & male	7	15	21%	8.90	−.31	.24	.220	
Non gender targeted & male[a]	8	19	27%	13.80	.35	.26	.208	
Intervention duration								
Intercept (1 week)	19	67		5.74	.27	.15	.119	
Duration				2.96	.01	.01	.296	
Outcome type								
Attitudes	5	6	8%	3.99	.52	.23	.092	REF
Affective	7	16	23%	7.97	−.28	.27	.322	
Motivational	9	30	42%	9.59	−.18	.27	.506	
Self-schema	9	19	27%	9.68	−.07	.26	.805	

Note: j represents the number of studies, *k* represents the number of effect sizes, % represents what percent of the effect sizes were at each of the various moderator categories, and REF indicates which level of the variable was used as the reference category (intercept). ***$p < .01$, **$p < .05$.

[a]Indicates that this combination was an interaction term in the meta-regression.

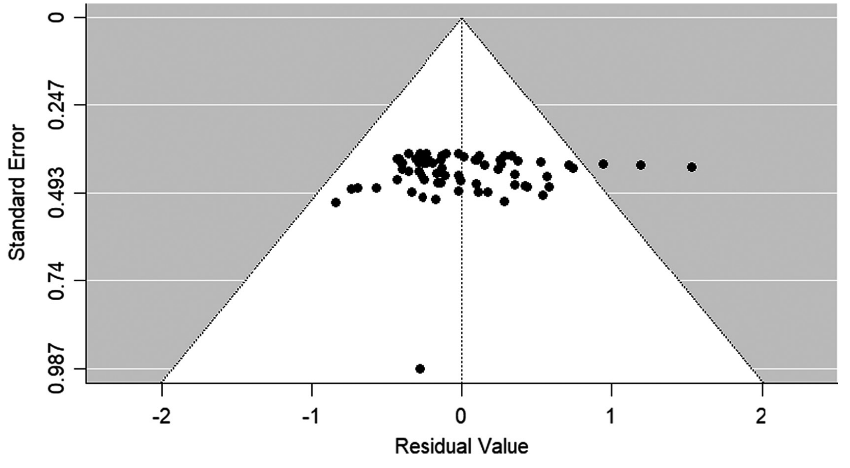

FIGURE 4 Funnel plot

female students is encouraging, as it demonstrates the efficacy of school-based interventions across genders. Descriptively, we found that the overall average effect size for females was larger than that for males; however, this difference in effect sizes was not statistically significant. While these results provide preliminary evidence that there may be a difference in how effective these interventions are based on student gender, more research is needed to determine if these differences in effects are significant, and if they hold true in a larger sample. One possible explanation for this may be that interventions function more effectively when levels of these motivational-affective factors are lower to start out with, which would explain the possibly larger average effect for females in subjects that are typically "female disadvantaged" where females have been shown to have lower levels of motivational-affective factors. Male students tend to have higher levels of these factors on average in subjects such as science or math, and therefore, while these interventions are still significantly and positively affecting them, they may not be as "in need" of support as females in these subjects.

The relevance of intervention characteristics

Descriptively, the results show that the intervention effects were slightly greater on average for students in primary school (Grades 1–5) and lower secondary school (Grades 6–8) than for students in upper secondary school (Grades 9–12). Although these differences were not statistically significant, the findings are in line with other studies, which have also found intervention effects to be stronger for students in childhood and early adolescence. For example, Lazowski and Hulleman (2016) found almost identical effects in a meta-analysis on motivation interventions in education, with the largest effect sizes for students in grades 6–8 ($d = .57$), followed closely by students in grades 1–5 ($d = .52$) and lastly, by grades 9–12 ($d = .42$). The literature suggests that these lower secondary and primary-school-aged groups tend to reap the most benefits from these interventions when it comes to motivational-affective factors and this stage of development might be the prime opportunity to target these factors in students (Juvonen, 2007; Wentzel & Wigfield, 2007).

This study was novel in that it included various types of interventions in order to gain a wide view of the current state of research, whereas most prior meta-analyses on school-based interventions have focused on one specific intervention strategy, such as digital media (Hillmayr et al., 2020) or problem-based learning (Batdi, 2014). We differentiated between psycho-social intervention approaches and instructional approaches. We found that both approaches had a descriptively larger effect sizes for females than for males; however, these gender differences were significant only for psycho-social approaches. The gender difference in effect sizes was particularly prominent for psycho-social interventions. Contrary to instructional interventions, psycho-social interventions are specifically designed to directly target motivational-affective factors in students (Walton & Wilson, 2018) and therefore could explain why the effect size was larger for females in stereotypically male subjects (with lower starting values). Although we did not test the overall differences between psycho-social and instructional interventions (independent from gender), strategies which aim directly for student motivation, attitude, emotions, or self-beliefs may be more effective at promoting these factors than other types of interventions.

Along the same line, our results showed larger descriptive effects for interventions that target females in mathematics or science than interventions that were not gender targeted. However, as our sample only included studies that were conducted in subjects where females were disadvantaged, the interventions likewise were targeted towards females only. Our results showed that in these female-targeted interventions, males had lower effect sizes than females, although not significantly different. All of this evidence points to the need for more interventions that directly target these factors in students, as well as interventions that consider which gender might be disadvantaged in a certain subject.

The finding that intervention duration did not significantly moderate intervention effects fits with prior research, as there seem to be mixed results on if and how the length of an intervention moderates its effects. While some studies have found small effects for intervention duration (Hattie et al., 1996; Slavin & Lake, 2009), other meta-analyses on interventions have also not found any significant effects of intervention duration (van Eerde & Klingsieck, 2018). However, given the relatively small sample size of our meta-analysis, and that we treated intervention duration as a continuous variable, it is possible that we did not have enough power to detect any effects that might be present. More research is needed to determine if the length of an intervention is overall a moderating variable of the intervention effects, and if there is an effect of length, whether or not it differs depending on student gender.

There were no significant differences between the type of motivational-affective outcome measured. Descriptively, the biggest effect size could be seen for when attitudes were the measured outcome. This may be due to the broader definition of attitudes, as compared to the other categories of outcomes measured. Measures of attitude usually encompass many various aspects such as value, self-efficacy, beliefs, and relevance (Thurstone, 1970), and therefore capture a wider range of student factors.

Limitations & future research

While this study provides a crucial first look at gender-specific effects of school-based interventions, there are certain limitations that must be considered when interpreting the results. Out of 171 potentially eligible articles, we excluded 35 due to missing data or information. This was due to two main issues. First, while many studies were eligible for inclusion in this meta-analysis based on the characteristics of the interventions themselves, many of them could not be included because they did not evaluate gender-specific effects, and therefore, the data necessary for calculating effect sizes for the different genders were not available. This highlights the need for future research to evaluate and report not only the overall effects of interventions but also the gender-specific effects, especially in subjects where there is typically a gender gap in these factors. Second, we excluded a number of studies because they did not include a control group or did not include any pre-test measures, and therefore did not fulfil our strict criteria for study design. This illustrates the greater need for high-quality intervention research in education, with stricter adherence to standards for effective intervention studies, as robust meta-analytic results rely on well-powered and well-designed primary studies.

These issues with missing information from primary studies resulted in a relatively smaller sample size, leading to a number of methodological limitations and changes to our original study design. Firstly, while we originally planned to include all moderators into one model, our final sample size did not allow us to robustly test this model. We therefore evaluated each moderator individually, which did not allow us to control for any possible confounding interactions between moderators. It is possible that relationships between moderators could have had confounding effects on the results of our analyses. More high-quality intervention studies on gender differences in interventions would allow future meta-analytic studies to control for these possible interactions, and gain a closer look at how these moderators might relate to each other.

Due to the limited sample, we also were not able to examine interventions conducted in subjects where male students are typically disadvantaged. This highlights the need for more intervention research in subjects where male students are stereotypically disadvantaged. A surge of studies have focused on increasing female students interest, performance, and participation in STEM subjects in recent years (Kanny et al., 2014). While this is undoubtedly an important topic, it is equally as important to continue research in subject where male students are disadvantaged as well. Reading, writing, and language arts is a subject where boys have consistently displayed lower levels of interest, self-efficacy, and motivation (OECD, 2019; Retelsdorf et al., 2015). Increasing intervention research in these subjects is also a crucial step in closing the gender gap between all students.

Another limitation that must be considered is the quality of the included studies. We had certain quality requirements for studies to be included, namely that they use a pre-/post-test design and include a control group. We also assessed study quality through multiple additional criteria, which were study design (experimental vs. quasi-experimental), instrument reliability (high vs. low), instrument source (established vs. self-developed), and equivalence at baseline. While almost all studies used highly reliable, established instruments to measure their outcomes, 18 of the 71 included effect sizes did not meet the recommended standards of the What Works Clearinghouse for baseline equivalence. Due to our relatively small sample size and the dependency of our data (males vs. females), we chose not to exclude these effect sizes in order to include as many primary studies as possible in an area with an already limited number of intervention studies. However, this is a point that must be taken into account when interpreting our results. Additionally, we chose to include studies that used non-random assignment when selecting treatment and control groups. In educational research, it is often difficult to use completely random assignment when conducting studies in schools or classrooms, as separating already existing groups of students is usually impractical and, at times, unethical (Gopalan et al., 2020). Due to this, many studies conducted in schools will often use pre-existing classes as a treatment and control group, making this method of group assignment quite common throughout the literature. In order to avoid excluding any potentially relevant interventions, we chose to include these studies. However, using pre-existing classes for group assignment could be a confounding factor in these studies. Even though we statistically adjusted for these clustered assignments, future research should aim to examine studies that only use random assignment in order to eliminate any potentially confounding effects.

Lastly, we had a great degree of statistical heterogeneity among primary studies. Although the moderator analyses revealed several interesting descriptive trends regarding various characteristics of school-based interventions, none of them significantly influenced the intervention effects, and we were not able to completely explain the amount of heterogeneity. There are numerous contextual characteristics that were not taken into account in this study that may have played a role in the varying effect sizes. For example, a number of other factors have been shown to be connected to both gender stereotypes, as well as motivational-affective factors, such as high-achieving vs. low-achieving students, socio-economic status and cultural background (Dietrich et al., 2013; Guo et al., 2015; Rowley et al., 2007; Rozek et al., 2015). These features were not examined in the current study due to the lack of sufficient information from primary studies and low statistical power, but they are important variables for future researchers to consider including when examining the gender-specific effects of school-based interventions.

Implications and conclusion

This meta-analysis provides a comprehensive overview of the interventions on student motivational-affective factors and their effectiveness, and sheds light on the need to investigate what strategies are most effective in promoting and strengthening these factors in student in order to gain a deeper empirical understanding. This provides an important step forward in continuing to investigate ways in which schools and teachers can combat gaps that arise between male and female students in motivation, interest, self-efficacy, enjoyment, and engagement. School-based interventions are clearly a promising method for promoting the motivational-affective factors of school-aged children and adolescents. We defined school-based interventions as any method used in a school context, which is different from regular instruction. This includes not only in-class interventions but also novel teaching methods, summer school programmes, or school-organized workshops. A variety of possibilities are there for implementing these types of interventions into a school curriculum. The results of this study show that these interventions seem to have positive effects, and that researchers should continue to pursue investigations into these interventions in order to gain a better understanding of what interventional strategies are most effective for promoting student motivational-affective factors overall. Future studies should build on this current work, using strong theoretical frameworks with regard to the development of student motivation and affect to design and test school-based interventions.

This study also provides a comprehensive overview of high-quality intervention studies to date that have evaluated the gender-specific effects of school-based interventions. Our findings provide possible evidence that gender may play a role in the effectiveness of a given intervention. This is something that is evident in a number of primary studies. For example, the study by Falco et al. (2008) designed an intervention to improve student self-efficacy beliefs through fostering various skills such as time management and goal-setting, based on social-cognitive and expectancy-value theories. This primary study found that while all students in the experimental group developed more positive attitudes compared to the control, the gains for female students were significantly higher. Researchers should continue to investigate how these interventions affect male and female students differently, and conduct more studies on promising strategies to build a strong evidence base for practitioners and policy-makers, who can then make evidence-based recommendations for best classroom practices to combat gender differences between students.

This meta-analysis also identified contextual variables of these interventions such as intervention method, student grade level, and school subject that might play a role in how effective these interventions are for students. Future research should investigate this more deeply to determine how salient these effects may be. Continued research on this topic will help to create educational settings that are more inclusive and assist all students equally in achieving their full potential, regardless of gender.

ACKNOWLEDGEMENTS

We would like to thank the student research assistants who helped with the data collection and preparation for this study. The project on which this report is based was funded by the German Federal Ministry of Education and Research and the Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Federal States in the Federal Republic of Germany [ZIB2022].

CONFLICT OF INTEREST

All authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Kaley Lesperance: Conceptualization; Data curation; Investigation; Methodology; Project administration; Resources; Writing – original draft; Writing – review & editing. **Sarah Hofer:** Conceptualization; Investigation; Methodology; Project administration; Supervision; Writing – original draft; Writing – review & editing. **Jan Retelsdorf:** Conceptualization; Writing – review & editing. **Doris Holzberger:** Conceptualization; Methodology; Project administration; Supervision; Visualization; Writing – review & editing.

DATA AVAILABILITY STATEMENT

The data supporting this meta-analysis will be from previously reported studies and datasets, which have been cited. The processed data will be made openly available upon publishing of the study.

ORCID

Kaley Lesperance https://orcid.org/0000-0002-0669-2744
Sarah Hofer https://orcid.org/0000-0001-7267-9356
Jan Retelsdorf https://orcid.org/0000-0002-9892-9613
Doris Holzberger https://orcid.org/0000-0001-6535-289X

REFERENCES

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

*Abed, O. H. (2016). Drama-based science teaching and its effect on students' understanding of scientific concepts and their attitudes towards science learning. *International Education Studies*, *9*(10), 163–173. https://doi.org/10.5539/ies.v9n10p163

*Akcay, H., Yager, R. E., Iskander, S. M., & Turgut, H. (2010). Change in student beliefs about attitudes toward science in grades 6-9. In *Asia-pacific forum on science learning and teaching* (Vol. *11*, No. 1, pp. 1–18). The Education University of Hong Kong, Department of Science and Environmental Studies.

*Al-Balushi, S. M., & Al-Aamri, S. S. (2014). The effect of environmental science projects on students' environmental knowledge and science attitudes. *International Research in Geographical and Environmental Education*, *23*(3), 213–227. https://doi.org/10.1080/10382046.2014.927167

*Alghamdi, A. (2017). *Impact of jigsaw on the achievement and attitudes of saudi arabian male high school science students* [Doctoral dissertation, University of Akron].

Annetta, L. A. (2008). Video games in education: Why they should be used and how they are being used. *Theory Into Practice*, *47*, 229–239. https://doi.org/10.1080/00405840802153940

*Argaw, A. S., Haile, B. B., Ayalew, B. T., & Kuma, S. G. (2016). The effect of problem based learning (PBL) instruction on students' motivation and problem solving skills of physics. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(3), 857–871. https://doi.org/10.12973/eurasia.2017.00647a

Batdi, V. (2014). The effects of a problem based learning approach on students' attitude levels: A meta-analysis. *Educational Research and Reviews*, *9*(9), 272–276. https://doi.org/10.5897/ERR2014.1771

*Batton, M. (2010). *The effect of cooperative groups on math anxiety* [Doctoral dissertation, Walden University]. Walden Dissertations and Doctoral Studies.

Begg, C. B., & Mazumdar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, *50*(4), 1088. https://doi.org/10.2307/2533446

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist*, *34*, 139–153. https://doi.org/10.1207/s15326985ep3403_1

Borenstein, M., Higgins, J. P., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. https://doi.org/10.1002/jrsm.1230

Bruder, R., & Prescott, A. (2013). Research evidence on the benefits of IBL. *ZDM Mathematics Education*, *45*, 811–822. https://doi.org/10.1007/s11858-013-0542-2

Burke, R., & Mattis, M. (2007). *Women and minorities in science, technology, engineering and mathematics: Upping the numbers*. Edward Elgar Publishing. https://doi.org/10.4337/9781847206879

*Cady, D., & Terrell, S. R. (2008). The effect of the integration of computing technology in a science curriculum on female students' self-efficacy attitudes. *Journal of Educational Technology Systems*, *36*(3), 277–286. https://doi.org/10.2190/ET.36.3.d

Casad, B. J., Hale, P., & Wachs, F. L. (2015). Parent-child math anxiety and math-gender stereotypes predict adolescents' math education outcomes. *Frontiers in Psychology*, *6*, 1597. https://doi.org/10.3389/fpsyg.2015.01597

*Chen, H. T., Wang, H. H., Lu, Y. Y., & Hong, Z. R. (2019). Bridging the gender gap of children's engagement in learning science and argumentation through a modified argument-driven inquiry. *International Journal of Science and Mathematics Education*, *17*(4), 635–655. https://doi.org/10.1007/s10763-018-9896-9

*Chiu, M. S. (2011). Effects of a women-in-sciences/men-in-humanities intervention on Taiwanese adolescents' attitudes towards learning science. *Asia-Pacific Education Researcher (De La Salle University Manila)*, *20*(2), 322–335.

Christensen, R. (2002). Effects of technology integration education on the attitudes of teachers and students. *Journal of Research on Technology in Education*, *34*, 411–433. https://doi.org/10.1080/15391523.2002.10782359

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org//10.1037/h0026256

Dasgupta, N., & Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 21–29. https://doi.org/10.1177/2372732214549471

de Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, *84*, 509–545. https://doi.org/10.3102/0034654314540006

Dietrich, J., Schnabel, K., Ortner, T. M., Eagly, A., Retamero-Garcia, R., Kröger, L., & Holst, E. (2013). Internalized gender stereotypes vary across socioeconomic indicators. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2273466

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*, 231–264. https://doi.org/10.1007/s11409-008-9029-x

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, *98*, 382–393. https://doi.org/10.1037/0022-0663.98.2.382

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eccles, J. (1994). Understanding women's educational and occupational choices: Applying the Eccles et al model of achievement-related choices. *Psychology of Women Quarterly*, *18*, 585–609. https://doi.org/10.1111/j.1471-6402.1994.tb01049.x

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, *64*, 830–847. https://doi.org/10.2307/1131221

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127. https://doi.org/10.1037/a0018053

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*, 156–167. https://doi.org/10.1016/j.compedu.2013.02.019

*Fabian, K., Topping, K. J., & Barron, I. G. (2018). Using mobile technologies for mathematics: Effects on student attitudes and achievement. *Educational Technology Research and Development*, *66*(5), 1119–1139. https://doi.org/10.1007/s11423-018-9580-3

*Falco, L. D., Crethar, H., & Bauman, S. (2008). Skill-builders: Improving middle school students' self-beliefs for learning mathematics. *Professional School Counseling*, *11*(4), 229–235. https://doi.org/10.5330/PSC.n.2010-11.229

*Falco, L. D., & Summers, J. J. (2019). Improving career decision self-efficacy and STEM self-efficacy in high school girls: Evaluation of an intervention. *Journal of Career Development*, *46*(1), 62–76. https://doi.org/10.1177/0894845317721651

*Fennema, E., Pedro, J. D., Wolleat, P. L., & Becker, A. D. (1981). Increasing women's participation in mathematics: An intervention study. *Journal for Research in Mathematics Education*, *12*(1), 3–14. https://doi.org/10.2307/748654

Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. *arXiv* [Preprint]. https://doi.org/10.48550/arXiv.1503.02220

Goetz, T., Frenzel, A. C., Pekrun, R., & Hall, N. C. (2006). The domain specificity of academic emotional experiences. *The Journal of Experimental Education*, *75*, 5–29. https://doi.org/10.3200/JEXE.75.1.5-29

Goetz, T., & Hall, N. C. (2013). Emotion and achievement in the classroom. In J. Hattie, & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 192–195). Routledge.

Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, *44*(1), 218–243. https://doi.org/10.3102/0091732X20903302

Green, J., Martin, A. J., & Marsh, H. W. (2007). Motivation and engagement in English, mathematics and science high school subjects: Towards an understanding of multidimensional domain specificity. *Learning and Individual Differences*, *17*, 269–279. https://doi.org/10.1016/j.lindif.2006.12.003

Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*, 153–166. https://doi.org/10.1007/s11199-011-9996-2

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, *51*(8), 1163–1176. https://doi.org/10.1037/a0039440

Gutman, L. M., & Schoon, I. (2014). *The impact of non-cognitive skills on outcomes for young people* (pp. 1–5). Education Endowment Foundation.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, *66*(2), 99–136. https://doi.org/10.3102/00346543066002099

Heckman, J., & Kautz, T. (2013). *Fostering and measuring skills: Interventions that improve character and cognition (No. w19656)*. National Bureau of Economic Research. https://doi.org/10.3386/w19656

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*, 411–482. https://doi.org/10.1086/504455

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*, 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://doi.org/10.3102/0162373707299706

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. https://doi.org/10.1037/1082-989X.3.4.486

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, *26*, 276–295. https://doi.org/10.1016/j.appdev.2005.02.007

Heyder, A., Steinmayr, R., & Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept? *Frontiers in Education*, *4*, 34. https://doi.org/10.3389/feduc.2019.00034

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, *153*, 103897. https://doi.org/10.1016/j.compedu.2020.103897

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*, 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hsu, H.-Y., Chen, S.-H., Yu, H.-Y., & Lou, J.-H. (2010). Job stress, achievement motivation and occupational burnout among male nurses. *Journal of Advanced Nursing*, *66*, 1592–1601. https://doi.org/10.1111/j.1365-2648.2010.05323.x

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*, 880–895. https://doi.org/10.1037/a0019506

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *British Medical Journal Open*, *6*(7), e010247. https://doi.org/10.1136/bmjopen-2015-010247

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527. https://doi.org/10.1111/1467-8624.00421

Jones, M. G., Howe, A., & Rua, M. J. (2000). Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science Education*, *84*, 180–192.

Juvonen, J. (2007). Reforming middle schools: Focus on continuity, social connectedness, and engagement. *Educational Psychologist*, *42*(4), 197–208. https://doi.org/10.1080/00461520701621046

Kanny, M. A., Sax, L. J., & Riggers-Piehl, T. A. (2014). Investigating forty years of STEM research: How explanations for the gender gap have evolved over time. *Journal of Women and Minorities in Science and Engineering*, *20*(2), 127–148. https://doi.org/10.1615/JWomenMinorScienEng.2014007246

*Kerneža, M., & Košir, K. (2016). Comics as a literary-didactic method and their use for reducing gender differences in reading literacy at the primary level of education. *Center for Educational Policy Studies Journal*, *6*(2), 125–149. https://doi.org/10.26529/cepsj.91

Kim, C., & Pekrun, R. (2014). Emotions and motivation in learning and performance. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 65–75). Springer. https://doi.org/10.1007/978-1-4614-3185-5_6

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal of Research in Mathematics Education*, *32*, 448–470. https://doi.org/10.2307/749801

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, *31*, 486–490. https://doi.org/10.1016/j.sbspro.2011.12.091

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Leaper, C. (2015). Gender and social-cognitive development. In R. M. Lerner, L. S. Liben, & U. Muller (Eds.), *Handbook of child psychology and developmental science* (pp. 1–48). American Cancer Society. https://doi.org/10.1002/9781118963418.childpsy219

Lieberman, D. A., Bates, C. H., & So, J. (2009). Young children's learning with digital media. *Computers in the Schools*, *26*, 271–283. https://doi.org/10.1080/07380560903360194

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research*, *37*, 888–902. https://doi.org/10.1016/j.ssresearch.2008.03.004

Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading*, *32*, 199–214. https://doi.org/10.1111/j.1467-9817.2008.01389.x

Marinak, B. A., & Gambrell, L. B. (2010). Reading motivation: Exploring the elementary gender gap. *Literacy Research and Instruction*, *49*, 129–141. https://doi.org/10.1080/19388070902803795

Marsh, H. W., Martin, A. J., & Debus, R. (2001). Individual differences in verbal and math self-perceptions: One factor, two factors, or does it depend on the construct. *International Perspectives on Individual Differences*, *2*, 149–170.

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, *52*, 1119–1134. https://doi.org/10.2307/1129498

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology*, *160*, 92–106. https://doi.org/10.1016/j.jecp.2017.03.013

Mavridis, A., Katmada, A., & Tsiatsos, T. (2017). Impact of online flexible games on students' attitude towards mathematics. *Educational Technology Research and Development*, *65*(6), 1451–1470. https://doi.org/10.1007/s11423-017-9522-5

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749. https://doi.org/10.1177/1745691616662243

Miller, P. H., Slawinski Blessing, J., & Schwartz, S. (2006). Gender differences in high-school students' views about science. *International Journal of Science Education*, *28*(4), 363–381. https://doi.org/10.1080/09500690500277664

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research*, *90*(3), 376–419. https://doi.org/10.3102/0034654320919354

Morgenroth, T., Ryan, M. K., & Peters, K. (2015). The motivational theory of role modeling: How role models influence role aspirants' goals. *Review of General Psychology*, *19*, 465–483. https://doi.org/10.1037/gpr0000059

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*, 364–386. https://doi.org/10.1177/1094428106291059

Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—The role of teachers' gender stereotypes. *Contemporary Educational Psychology*, *54*, 212–220. https://doi.org/10.1016/j.cedpsych.2018.06.012

Muntoni, F., Wagner, J., & Retelsdorf, J. (2020). Beware of stereotypes: Are classmates' stereotypes associated with students' reading outcomes? *Child Development*, *92*(1), 189–204. https://doi.org/10.1111/cdev.13359

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

OECD. (2013). Mathematics self-beliefs and participation in mathematics-related activities. In *OECD, PISA 2012 results: Ready to learn* (Vol. *III*, pp. 87–112). OECD. https://doi.org/10.1787/9789264201170-8-en

OECD. (2014). *Indicator D5: Who are the teachers (Education at a Glance 2014: OECD Indicators)*. OECD Publishing. https://doi.org/10.1787/eag-2018-en

OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education* (Vol. *1*). OECD. https://doi.org/10.1787/9789264266490-en

OECD. (2019). *PISA 2018 results (Volume II): Where all students can succeed*. OECD. https://doi.org/10.1787/b5fd1b8f-en

*Orabuchi, N. (2013). *Effects of online visual and interactive technological tool (OVITT) on early adolescent students' mathematics performance, math anxiety and attitudes toward math* [Doctoral dissertation, Texas Woman's University]. ProQuest Dissertations Publishing.

Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. In A. M. Gallagher, & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 294–315). Cambridge University Press. https://doi.org/10.18848/2327-7920/CGP/v23i03/1-9

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*, 29–48. https://doi.org/10.1080/01443410.2013.797339

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. https://doi.org/10.1016/j.cedpsych.2010.10.002

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. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, *88*, 1653–1670. https://doi.org/10.1111/cdev.12704

Pintrich, P. R. (2003). Motivation and classroom learning. In W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 103–122). Wiley. https://doi.org/10.1002/0471264385.wei0706

Plante, I., de la Sablonnière, R., Aronson, J. M., & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology*, *38*, 225–235. https://doi.org/10.1016/j.cedpsych.2013.03.004

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). "Michael can't read!" Teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology*, *107*(1), 186–194. https://doi.org/10.1037/a0037107

Rosen, J. A., Glennie, E. J., Dalton, B. W., Lennon, J. M., & Bozick, R. N. (2010). *Noncognitive skills in the classroom: New perspectives on educational research*. RTI International.

Rowley, S. J., Kurtz-Costes, B., Mistry, R., & Feagans, L. (2007). Social status as a predictor of race and gender stereotypes in late childhood and early adolescence. *Social Development*, *16*, 150–168. https://doi.org/10.1111/j.1467-9507.2007.00376.x

Rozek, C. S., Hyde, J. S., Svoboda, R. C., Hulleman, C. S., & Harackiewicz, J. M. (2015). Gender differences in the effects of a utility-value intervention to help parents motivate adolescents in mathematics and science. *Journal of Educational Psychology*, *107*(1), 195–206. https://doi.org/10.1037/a0036981

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. https://doi.org/10.1002/sce.21007

Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. Organisation for Economic Co-operation and Development.

Schmenk, B. (2004). Language learning: A feminine domain? The role of stereotyping in constructing gendered learner identities. *TESOL Quarterly*, *38*, 514–524. https://doi.org/10.2307/3588352

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, *66*, 175–183. https://doi.org/10.1007/s11199-011-0051-0

*Shin, D. D., Lee, M., Ha, J. E., Park, J. H., Ahn, H. S., Son, E., Chung, Y., & Bong, M. (2019). Science for all: Boosting the science motivation of elementary school students with utility value intervention. *Learning and Instruction*, *60*, 104–116. https://doi.org/10.1016/j.learninstruc.2018.12.003

Slavin, R. E. (2011). Instruction based on cooperative learning. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 344–360). Routledge. https://doi.org/10.4324/9780203839089.ch17

Slavin, R. E., & Lake, C. (2009). *Effective educational programs: Meta-findings from the best evidence encyclopedia*. Annual Meetings of the Society for Research on Educational Effectiveness, 8.

*Soyibo, K., & Hudson, A. (2000). Effects of computer-assisted instruction (CAI) on 11th graders' attitudes to biology and CAI and understanding of reproduction in plants and animals. *Research in Science & Technological Education*, *18*(2), 191–199. https://doi.org/10.1080/713694977

*Starkey, P. (2013). *The effects of digital games on middle school students' mathematical achievement* [Doctoral dissertation, Lehigh University]. ProQuest Dissertations Publishing.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*(1), 80–90. https://doi.org/10.1016/j.lindif.2008.05.004

*Sung, Y. T., Shih, P. C., & Chang, K. E. (2015). The effects of 3D-representation instruction on composite-solid surface-area learning for elementary school students. *Instructional Science*, *43*(1), 115–145. https://doi.org/10.1007/s11251-014-9331-8

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. https://doi.org/10.1007/s40865-016-0026-5

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, *88*(4), 1156–1171. https://doi.org/10.1111/cdev.12864

Thurstone, L. L. (1970). Attitudes can be measured. In G. F. Summers (Ed.), *Attitude measurement* (pp. 127–141). Rand McNally.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, *92*(1), 144–151. https://doi.org/10.1037/0022-0663.92.1.144

Uzunboylu, H., & Karagozlu, D. (2015). Flipped classroom: A review of recent literature. *World Journal on Educational Technology*, *7*, 142–147. https://doi.org/10.18844/wjet.v7i2.46

*van der Meij, H., van der Meij, J., & Harmsen, R. (2015). Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. *Educational Technology Research and Development*, *63*(3), 381–403. https://doi.org/10.1007/s11423-015-9378-5

van Eerde, W., & Klingsieck, K. B. (2018). Overcoming procrastination? A meta-analysis of intervention studies. *Educational Research Review*, *25*, 73–85. https://doi.org/10.1016/j.edurev.2018.09.002

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. https://doi.org/10.1007/BF02294384

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, *23*, 73–82. https://doi.org/10.1177/0963721413512856

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, *125*(5), 617. https://doi.org/10.1037/rev0000115

Wang, S.-K., & Reeves, T. C. (2007). The effects of a web-based learning environment on student motivation in a high school earth science course. *Educational Technology Research and Development*, *55*, 169–192. https://doi.org/10.1007/s11423-006-0638-2

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, *32*, 387–398. https://doi.org/10.1002/tea.3660320407

Wentzel, K. R., & Wigfield, A. (2007). Motivational interventions that work: Themes and remaining issues. *Educational Psychologist*, *42*(4), 261–271. https://doi.org/10.1080/00461520701621103

What Works Clearinghouse. (2020). *Standards handbook* (Version 4.1). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Wigfield, A., Battle, A., Keller, L. B., & Eccles, J. S. (2002). Sex differences in motivation, self-concept, career aspiration, and career choice: Implications for cognitive development. In A. V. McGillicuddy, & R. De Lisi (Eds.), *Biology, society, and behavior: The development of sex differences in cognition* (Vol. *21*, pp. 93–124). Ablex.

Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem: Change across elementary and middle School. *Journal of Early Adolescence*, *14*, 107–138. https://doi.org/10.1177/027243169401400203

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading: Domain specificity and instructional influences. *Journal of Educational Research*, *97*, 299–310. https://doi.org/10.3200/JOER.97.6.299-310

*Zhao, F., Zhang, Y., Alterman, V., Zhang, B., & Yu, G. (2018). Can math-gender stereotypes be reduced? A theory-based intervention program with adolescent girls. *Current Psychology*, *37*(3), 612–624. https://doi.org/10.1007/s12144-016-9543-y

How to cite this article: Lesperance, K., Hofer, S., Retelsdorf, J., & Holzberger, D. (2022). Reducing gender differences in student motivational-affective factors: A meta-analysis of school-based interventions. *British Journal of Educational Psychology*, *92*, 1502–1536. https://doi.org/10.1111/bjep.12512

APPENDIX A

Search syntax for ERIC

(SU (students) AND SU (education* OR school OR classroom OR parents OR teachers) AND SU (female OR male OR gender differences OR sex stereotypes OR sex fairness OR gender bias OR gender issues OR sex role) AND AB ("intervention*" OR "mentor*" OR "role model*" OR "training*" OR "program*" OR "instruction*" OR "strategy*" OR "support*" OR "outreach*" OR "teaching*" OR "experiment*" OR "control group*") AND SU (interest OR self concept OR self esteem OR self efficacy OR motivation OR attribution theory OR stereotype OR career choice OR attitude OR beliefs OR values OR learner engagement OR participation OR satisfaction)) NOT SU (higher education OR universities OR college*)

Search syntax for PsycINFO

(SU (students) AND SU (education* OR school OR classroom OR parents OR teachers) AND SU ("human females" OR "human males" OR sex differences OR gender gap OR gender equality OR sex role) AND AB ("intervention*" OR "mentor*" OR "role model*" OR "training" OR "program" OR "instruction" OR "strategy*" OR "support" OR "outreach" OR "teaching" OR "experiment*" OR "control group*") AND SU (belonging OR interest OR self concept OR self-confidence OR self-esteem OR self-efficacy OR motivation OR attribution OR occupational choice OR attitude OR occupational preference OR values OR student engagement OR cognitive appraisal OR participation OR expectations)) NOT SU (higher education OR college*)

Search syntax for Web of Science

((((TS=(student*) AND TS=("education*" OR "school*" OR "classroom*" OR "teacher*" OR "parent*") AND TS=("gender gap*" OR "gender difference*" OR "gender stereotype*" OR "gender equality*" OR "gender bias*" OR "gender specific*" OR "sex difference*" OR "sex stereotype*" OR "sex role*") AND TS=("intervention*" OR "mentor*" OR "role model*" OR "training*" OR "program*" OR "instruction*" OR "strategy*" OR "support*" OR "outreach*" OR "teaching*" OR "experiment*" OR "control group*") AND TS=("interest*" OR "belonging*" OR "self concept*" OR "self efficacy*" OR "self confidence*" OR "self esteem*" OR "motivation*" OR "attribution*" OR "stereotype*" OR "career choice*" OR "attitude*" OR "belief*" OR "value*" OR "engagement*" OR "participation*" OR "expectation*")))))

APPENDIX B

Subgroup analyses

TABLE A1 Meta-analytic subgroup model for females in science

Predictor	Estimate	*SE*	95% CI Lower	95% CI Upper	*df*	*p*-value
Intercept	.389	.135	.052	.725	5.64	.0303**

Note: Model estimated using robust variance estimation to correct for dependent effect sizes. $^{**}p < .05$.

TABLE A2 Meta-analytic subgroup model for males in science

Predictor	Estimate	*SE*	95% CI Lower	95% CI Upper	*df*	*p*-value
Intercept	.247	.187	−.211	.705	5.91	.234

Note: Model estimated using robust variance estimation to correct for dependent effect sizes.

TABLE A3 Meta-analytic subgroup model for females in math

			95% CI			
Predictor	**Estimate**	***SE***	**Lower**	**Upper**	***df***	***p*-value**
Intercept	.379	.163	.001	.756	7.85	.049*

Note: Model estimated using robust variance estimation to correct for dependent effect sizes. *p < .10.

TABLE A4 Meta-analytic subgroup model for males in math

			95% CI			
Predictor	**Estimate**	***SE***	**Lower**	**Upper**	***df***	***p*-value**
Intercept	.227	.092	−.004	.457	5.51	.053*

Note: Model estimated using robust variance estimation to correct for dependent effect sizes. *p < .10.

APPENDIX C

TABLE A5 Sensitivity analysis for baseline equivalency

			95% CI			
Predictor	**Estimate**	***SE***	**Lower**	**Upper**	***df***	***p*-value**
Intercept	−.127	.069	−.278	.0.024	11.3	.0909*
Unequal baseline	.526	.080	.337	.716	6.9	.0003***

Note: Model estimated using robust variance estimation to correct for dependent effect sizes. * p < .10, ***p < .01.