



# Effects of games in STEM education: a meta-analysis on the moderating role of student background characteristics

Michaela Arztmann, Lisette Hornstra, Johan Jeuring & Liesbeth Kester

To cite this article: Michaela Arztmann, Lisette Hornstra, Johan Jeuring & Liesbeth Kester (2022): Effects of games in STEM education: a meta-analysis on the moderating role of student background characteristics, Studies in Science Education, DOI: [10.1080/03057267.2022.2057732](https://doi.org/10.1080/03057267.2022.2057732)

To link to this article: <https://doi.org/10.1080/03057267.2022.2057732>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 29 Mar 2022.



[Submit your article to this journal](#)



Article views: 190



[View related articles](#)



[View Crossmark data](#)

# Effects of games in STEM education: a meta-analysis on the moderating role of student background characteristics

Michaela Arztmann <sup>a</sup>, Lisette Hornstra<sup>a</sup>, Johan Jeuring<sup>b</sup> and Liesbeth Kester <sup>a</sup>

<sup>a</sup>Department of Education, Utrecht University, Utrecht, The Netherlands; <sup>b</sup>Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

## ABSTRACT

Game-based learning has proven to be effective and is widely used in science education, but usually the heterogeneity of the student population is being overlooked. To examine the differential effects of game interventions in STEM (Science, Technology, Engineering and Mathematics) related subjects on diverse student groups, a meta-analysis has been conducted that included 39 studies that compared game-based learning interventions with traditional classrooms in primary and early secondary education. We found moderate positive effects on cognition ( $g = .67$ ), motivation ( $g = .51$ ), and behaviour ( $g = .93$ ). Additionally, substantial heterogeneity between studies was found. Moderator analyses indicated that primary school students achieve higher learning outcomes and experience game interventions as more motivating than secondary school students, whereas gender did not have any moderating effect. There were too few studies reporting information on the remaining moderators (socioeconomic status, migration background, and special educational needs) to include them in a multiple meta-regression model. Therefore, we assessed their role by separate moderator analyses, but these results need to be interpreted with caution. Additional descriptive analyses suggested that game-based learning may be less beneficial for students with low socioeconomic status compared to students with high socioeconomic status.

## ARTICLE HISTORY

Received 30 April 2021  
Accepted 27 January 2022

## KEYWORDS

Meta-analysis; game-based learning; motivation; STEM education; diversity

## Introduction

Technical advances lead to an increased demand of highly skilled employees in STEM (Science, Technology, Engineering and Mathematics) industries. Therefore, students need to acquire science literacy to ensure that they have the skillset to actively participate in today's society (European Commission, 2015; European Commission, 2021). Prior research suggests that many students are not very motivated for STEM education and for many students, their motivation to engage in STEM subjects decreases already in upper primary school grades (Renninger et al., 2015; Shin et al., 2019) which can lead to students choosing a different specialisation in later grades to avoid STEM subjects (Simpkins et al., 2006). Game-based learning is widely used in schools as an approach to motivate

students for science. It intends to satisfy students' psychological needs for autonomy, competence, and relatedness and thereby contribute to greater interest, motivation, engagement, and achievement (R.M. Ryan & Rigby, 2020).

Various literature reviews and meta-analyses have been conducted with regard to the effects of game-based learning on student outcomes. Overall, they concluded that game-based learning improves students' learning outcomes (Riopel et al., 2019; Vogel et al., 2006; Wouters et al., 2013), and contributes to students' motivation and engagement (Lamb et al., 2018; Zainuddin et al., 2020). Looking in particular at game-based learning in STEM education, results are mixed and vary from no significant effect on science learning (Wouters et al., 2013) to improved learning outcomes for all STEM subjects (Riopel et al., 2019; Tsai & Tsai, 2020). Although all these studies have provided valuable insights with regard to the overall effects and have shown that games have the potential to improve students' motivation and learning outcomes, more insight is needed regarding the differential effects of games for different groups of students. That is, even though games may, on average, yield positive outcomes, these outcomes may not be the same for all students. Students can differ from another with regard to their age and gender, but also regarding their background characteristics such as their socioeconomic status (SES), migration, or the presence of special educational needs (SEN). Given that the effects of games have been found to depend on prior knowledge (Lee et al., 2019; Shute et al., 2015) and game self-efficacy (Nietfeld, 2020), it could be that games are not equally effective for different student groups who may differ in prior knowledge or self-efficacy beliefs. These results align with prior research showing differential effectiveness for other educational interventions (e.g., Li et al., 2021). Likewise, it may be that the effects of games may also differ based on different student characteristics. Since games are widely used in education, it is important to gain insight into whether some groups of students benefit more from games than others. If games are less effective for certain groups of students, the wide implementation of games might actually contribute to the perpetuation of pre-existing differences between groups.

As classrooms become increasingly diverse (Fine-Davis & Faas, 2014; OECD, 2019), it is important to gain more insight into whether games are indeed beneficial for all groups of students. The present meta-analysis aims to extend previous game-based learning research by examining the potential moderating effects of different student characteristics, in early STEM education. We focus on early STEM education in primary and lower secondary education as early interventions could be particularly beneficial to counteract the widely reported decrease in motivation for STEM education (Renninger et al., 2015; Shin et al., 2019; Stewart et al., 2013).

## **Theoretical framework**

### ***Game-based learning***

Games are designed to be immersive and enjoyable activities (Kinzie & Joseph, 2008). They provide a safe environment for graceful failure (Homer et al., 2020; Kinzie & Joseph, 2008; Plass et al., 2015, 2020) and can offer adaptivity to engage each learner individually (Plass et al., 2015; Van Oostendorp et al., 2014). While there is no consensus on the exact definition of a game, we align with the

definition of Plass et al. (2020) who define game-based learning as learning tasks that are redesigned into games with a full range of game features to make it more interesting and effective for learning. Typical characteristics of games that are widely agreed upon are that games have clear rules of play, are challenging, and motivate a player to play (Plass et al., 2020). We chose to adopt this broad definition to avoid restricting the possible sample studies in this meta-analysis. Regarding STEM education, game-based learning can be used to demonstrate different science concepts and they enable students to work with phenomena that would be otherwise invisible, or they can explore locations that usually would be unreachable (Klopfer & Thompson, 2020). Moreover, games can provide learners with additional scaffolding and context, which may support the development of skills (i.e., problem solving) that are needed for STEM education (Homer et al., 2020; Klopfer & Thompson, 2020).

Game-based learning is believed to affect a wide range of outcomes. Plass et al. (2015) described four different theoretical perspectives for game-based learning research that refer to the game design as well as the outcomes that games are targeting: a *cognitive perspective*, a *motivational perspective*, an *affective perspective*, and a *sociocultural perspective*. The *cognitive perspective* of game-based learning focuses on the construction of mental models and as such is using theories describing cognitive aspects of learning. Research from this perspective typically includes cognitive outcome measures, such as learning outcomes. Research focusing on *motivational perspectives* of game-based learning is based on motivational theories such as self-determination theory (Ryan & Deci, 2000) or self-efficacy theory (Bandura, 1977) and focuses on games as a tool to engage and motivate players. The outcome measures associated with this perspective usually include measures of motivation and associated constructs such as interest and engagement. Studies based on affective theories, such as the control-value theory (Pekrun, 2006) focus on games as a means to create a more positive affective emotional experience for students. Outcome measurements mostly pertain to the emotions experienced by students (e.g., curiosity or boredom; Loderer et al., 2020). Research using *sociocultural perspectives* investigates how social interactions in game environments influence learning (Plass et al., 2015). As such, research could investigate the differences among single play, collaborative play, and competitive play or the identity construction within large game communities.

In a review on game-based learning in primary education, Hainey et al. (2016) distinguished four different categories of outcome variables that are most commonly examined in research, including (1) knowledge acquisition and content understanding, (2) affection and motivation, (3) perceptual and cognitive outcomes, and (4) behavioural change.

In the present study, we follow the categories of Hainey et al. (2016) for our outcome variables. Category (1) knowledge acquisition and content understanding and category (3) perceptual and cognitive outcomes were combined into one single category 'cognitive outcomes', since other frameworks (e.g., Plass et al., 2015) refer to cognition as a single category and due to the conceptual overlap between these two categories. Accordingly, we distinguish the following three types of outcomes of game-based learning:

- (1) *Cognitive outcomes*. This category aligns with Plass' cognitive perspective and combines Hainey's categories (1) knowledge acquisition and content understanding and (3) perceptual and cognitive outcomes, as these two categories contain relatively overlapping constructs that are hard to distinguish. Outcome variables within this category are for example, assessed learning outcomes, knowledge retention (Fokides, 2018b) or conceptual knowledge (Chee & Tan, 2012).
- (2) *Motivational outcomes*. In line with Hainey, affect and motivation were combined into one category *motivation*. We understand *motivation* as affect towards a task which can be called differently depending on the theory used (e.g., intrinsic motivation (Ryan & Deci, 2000), interest (Hidi & Renninger, 2006), self-efficacy (Bandura, 1977), enjoyment (Ryan & Deci, 2000)).
- (3) *Behavioural outcomes*. Aligning with Hainey, we added the category behaviour. This category includes various behavioural outcomes, such as self-reported strategy use, behavioural engagement, or effort (Fredricks et al., 2004).

Unlike the other three perspectives, the sociocultural perspective of Plass et al. (2020) does not specify a category of outcomes. Rather, it focuses on specific game features (individual versus collaborative), which is beyond the scope of the present study as the present study focuses on the moderating role of different student characteristics.

### **Background characteristics of students**

Games are an increasingly popular tool in schools and several meta-analyses have shown the beneficial effects of game-based learning (Lamb et al., 2018; Riopel et al., 2019; Tsai & Tsai, 2020; Vogel et al., 2006; Wouters et al., 2013). These meta-analyses have mostly focused on explaining heterogeneity between studies by moderators such as discipline (Riopel et al., 2019; Wouters et al., 2013), instructional methods (Riopel et al., 2019; Vogel et al., 2006; Wouters et al., 2013), and game features (Lamb et al., 2018; Riopel et al., 2019; Tsai & Tsai, 2020). Very limited attention has been paid to the potential differential effects of game-based learning for students with different background characteristics. As such, the question if and how the effects of games may differ across different student groups is largely unexplored.

Regular classrooms are typically characterised by heterogeneity. That is, students can differ in age, gender, migration background, SES, and in having special educational needs (e.g., ADHD, dyslexia). Some of these background characteristics of students are associated with a higher risk of school failure and early school leaving. According to the European Agency for Special Needs (2019) the following learner circumstances can have a negative impact on students' educational outcomes: (im)migration background, low SES, or special educational needs due to learning or behavioural disabilities. This finding is supported by results of other studies, where students with low SES, special educational needs and migration background score lower in STEM-related subjects than their peers (Betancur et al., 2018; Mullis et al., 2020; Garon-Carrier et al., 2018; Lie, Selcen Gucey, & Moore, 2019). Those differences can already be seen as early as in kindergarten. Children with disadvantageous learner circumstances show lower mathematical precursor abilities (Garon-Carrier et al., 2018), lower literacy, language, and self-regulatory skills (Verhaeghe

et al., 2018) than other children. Additionally, low parental income can negatively influence science learning opportunities inside and outside the home environment (e.g., museum visits), due to fewer resources available (Betancur et al., 2018).

Even though gender has not been classified as predictor that could negatively influence educational outcome (European Agency for Special Needs and Inclusive Education, 2019), there are gender differences in the domain of STEM. Women are still an under-represented group in STEM careers (Turner et al., 2019) and girls in general report lower science self-efficacy (Marriott et al., 2019). Lower self-efficacy can also be found in students from disadvantageous backgrounds with negative learning biographies. Students become discouraged over time with lower levels of hopefulness and high levels of anxiety, frustration, and distress (Bandura, 1977; Schunk & DiBenedetto, 2020). To reduce these negative emotions, students can adopt aversive behaviour as coping strategy (Bandura, 1977) which could explain loss of interest in the specific subject and in general the higher risk of early school dropout.

There are several reasons to assume that games may have differential effects for students with different background characteristics. Boys have been found to play digital games more frequently than girls (Homer et al., 2012; Hygen et al., 2019; Kinzie & Joseph, 2008) and boys' and girls' preferences for game characters and game type (e.g., strategic, problem-solving) can differ (Homer et al., 2012). Additionally, games seem to have the potential to motivate and to reach especially those students with negative learning biographies through their typical characteristics such as graceful failure, scaffolding or providing autonomy. At the same time, high achieving or gifted students could potentially get disengaged by game-based learning due to cognitive underload (Van Oostendorp et al., 2014). Studies have found that games could increase the social participation of at-risk students (Hanghøj et al., 2018), improve word reading skills for dyslectic children (Ronimus et al., 2019), have beneficial effects for training children with autism (Ke & Moon, 2018; Lu et al., 2018) and can increase prosocial behaviour among socioeconomically diverse students (Harrington & O'Connell, 2016). However, the lower availability of home resources (Mullis et al., 2020) could also indicate that not every student has the same access to technology and might not be used to playing digital games. Several studies have found differences in ICT literacy between students from different socioeconomic and ethnic backgrounds (Aesaert et al., 2015; Scherer & Siddiq, 2019; Volman et al., 2005), whereas gender differences are rather small (Gnambs, 2021; Scherer & Siddiq, 2019). Moreover, ICT skills are lower in younger students (Aesaert et al., 2015) yet increasing over time (Gnambs, 2021). These differences could be an indicator for possible inequalities in the effects of game-based learning as game-based learning research mostly focuses on digital games, which requires basic knowledge of the tool used.

Thus far, many overview studies have neglected to examine these student background characteristics as moderators of the effects of game-based learning and only age (Riopel et al., 2019; Tokac et al., 2018; Tsai & Tsai, 2020; Vogel et al., 2006; Wouters et al., 2013) and gender (Vogel et al., 2006) have been tested as moderators. Findings are not conclusive, as these overviews suggested less beneficial, though non-significant, effects for younger students (Riopel et al., 2019; Vogel et al., 2006; Wouters et al., 2013) as well as no differences between all age groups (Tokac et al., 2018; Tsai & Tsai, 2020). For gender, no significant differences were found (Vogel et al., 2006).

## The present study

The present study aims to extend previous game-based learning research by examining the potential moderating effects of different student characteristics in early STEM education. Therefore, we conducted three meta-analyses to examine the effect of game-based learning in STEM education on the outcome variables cognition, motivation, and behaviour. Our aim was to conduct moderator analyses to examine possible variations in effectiveness on game-based learning due to background characteristics of students and to identify whether this could possibly enlarge existing inequalities in classrooms. As such, we examined age (i.e., school level), gender, socioeconomic status, migration background, and special educational needs as potential moderators of the effects of game-based learning. The present meta-analysis could therefore be a valuable contribution both for research and practice, as it aims to give insight whether game-based learning works for every student equally and tries to highlight possible differences in effectiveness. This knowledge could help both researchers and educators to decide when and for whom to use game-based learning and whether they need to tailor their instruction when using game-based learning in classrooms. The scope of this study was to cover students in K1-8 and it answers the following research question: *How do effects of game-based learning differ for students with different characteristics (school level, gender, migration background, socio-economic status, special educational needs) in early STEM education?*

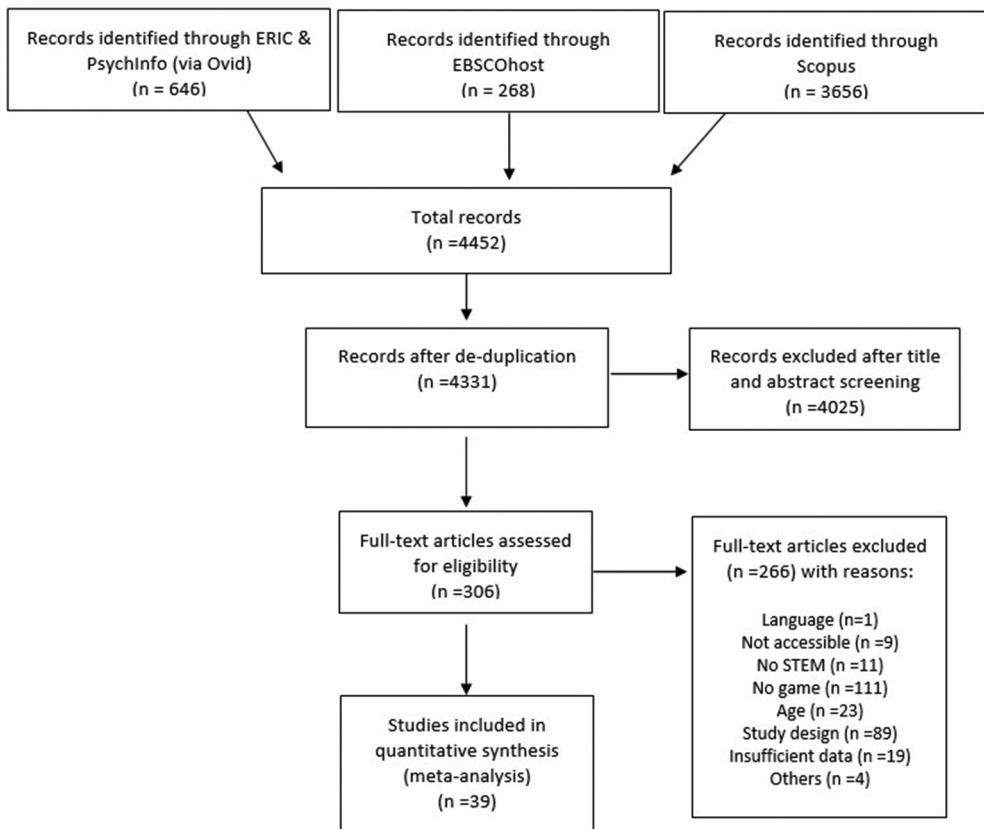
## Method

### Literature search

Published articles and unpublished research (e.g., dissertations) on game-based learning in early STEM education were searched in the databases ERIC and PsychInfo (via Ovid), Scopus, and EBSCOHost. We chose these databases, in consultation with an expert on search strategies, because these include both domain-specific as well as general databases. Since many relevant studies are interdisciplinary in nature, this combination of databases was expected to cover the entire range of relevant journals. Our aim was to discover all relevant studies, therefore our search terms included synonyms of game-based learning, STEM education, motivation, and learning outcomes and were used for free text search and subject headings. An example of a search string we used in PsychInfo (via Ovid) can be found in Appendix A. For a quality check of the search terms, we compared them to the search terms of previous meta-analyses and consulted a university librarian with expertise on search strategies for meta-analyses.

Additionally, we restricted the search to records published after 2000, with English as publication language, and to elementary and secondary school students (K1-8). The results of the literature search and the selection of articles are shown in the PRISMA diagram (Moher et al., 2009) in [Figure 1](#). The PRISMA diagram shows the numbers of articles found, included, and excluded at each step of this review.





**Figure 1.** Flow diagram of literature search and exclusion process.

### ***Inclusion of studies***

The first database search resulted in a total of 4452 papers. After deduplication, a total of 4331 papers were left for initial screening for which the platform Rayyan (Ouzzani et al., 2016) was used. In the initial screening phase, all papers were screened based on their titles and abstracts and papers that clearly did not fit the inclusion criteria presented in Table 1 (e.g., participants that were teachers instead of students, qualitative studies, no control group, simulations instead of a game) were excluded. In case the abstract was not providing enough information, the papers were included for the full-text screening. After the abstract screening, 306 papers were left for full-text assessment. Those articles were assessed for eligibility by applying the inclusion and exclusion criteria presented in Table 1. Aligning with the broad definition of ‘games’ we adopted, we included all studies which used the term ‘game’ for their intervention. Moreover, all included studies needed to investigate at least one of our defined outcome variables. As such, studies that examined, for example, knowledge retention, science literacy or conceptual understanding were included and coded for the outcome variable *cognition*.



**Table 1.** Inclusion and exclusion criteria.

| Criteria                | Inclusion   | Exclusion  |
|-------------------------|---|--|
| Language                | English   | Non-English language   |
| Subject                 | STEM  | Not STEM-related, e.g., geography  |
| Participants            | K 1–8 students  | preschool, high school, or university students   |
| Control Group           | Game vs non-game (e.g., traditional teaching)   | No control group, or control group is playing a different version of the game  |
| Availability            | Full-text is available  | Full-text is not available   |
| Game Definition         | Authors referring to the term ‘game’ regardless the technology, e.g., paper-based, AR, tablet | Studies that did not use the term ‘game’, e.g., online learning, flipped classroom, ‘simulation’ but not ‘simulation game’ |
| Dependent Variable      | Studies investigating cognitive, motivational, or behavioural outcome                         | Studies not including a dependent variable or investigating other dependent variables                                      |
| Statistical Information | Studies reporting sufficient statistical information to calculate the effect size             | Studies not reporting sufficient information to calculate the effect size  |

To establish reliability, at each screening step given in [Figure 1](#), 10% of all papers were screened independently by the first and second authors. or a trained research assistant. Disagreements were resolved through discussion (e.g., whether an outcome variable fitted within one of the three outcome categories). The interrater-reliability for the title and abstract screening was 84%, for the full-text screening 100%.

We excluded papers that were not in English, did not investigate the effects of games and where the dependent variable was not cognition, motivation or behaviour (for example, teachers’ perception, collaborative learning, or students’ perceptions of different game mechanics). Additionally, papers were excluded if they did not report enough data to calculate an effect size. In total, 39 papers were included. Two studies included data of independent groups that could be treated as individual studies. So, in total,  $k = 43$  studies were coded for this meta-analysis with a total of 60 effect sizes.

### ***Coding of study characteristics and moderator variables***

First, we coded basic information from the study, i.e., general study information (e.g., author, publication year), educational context features (e.g., academic domain), methodological characteristics (e.g., type of control group), and general sample characteristics (e.g., grade, gender, SES). Thereafter, information to calculate the effect size was coded, i.e., the outcome variable (cognition, motivation or behaviour) and statistical data needed to calculate the effect size (e.g., number of participants, mean scores). After two rounds of testing and adjusting the coding scheme, the first two authors coded 13% of the included papers ( $k = 6$ ) individually to check for interrater-reliability. For most variables, an interrater-reliability of Cohen’s  $k = 1.00$  could be reached; for the variables *study design* and *SES composition* differences in coding could be resolved after discussion.

Outcome variables were coded as cognition (e.g., learning outcomes), motivation (e.g., intrinsic motivation, interest, self-efficacy), or behaviour (e.g., self-reported strategy use, behavioural engagement, or effort). For background characteristics, the percentages of males, low SES students, migration background students, and special needs students in the sample were coded. Additionally, the grade, calculated as sample average, and two composition variables for SES and special educational needs were coded. Those

composition variables included detailed information about the SEN type and SES characteristics and were intended to help interpreting possible effects in the moderator analysis. Study characteristics (e.g., journal type, study design) were included to assess the effects of study quality on the outcomes. To code SES, we followed the definitions of the American Psychological Association (2007) and the TIMSS studies (Mullis et al., 2020). As such, low SES could be based either on *parental characteristics* (e.g., low-income level, low educational level, or low occupational level), *material and resources* (e.g., low support or few books at home, missing material), *geographic characteristics* (e.g., disadvantaged neighbourhood), or eligibility for *free lunch programs*. Special educational needs students were defined as students with either *learning disabilities* (e.g., dyslexia), *behavioural disabilities* (e.g., ADHD), or *gifted students*. Students with learning or behavioural disabilities were combined into one category, whereas gifted students were considered to be a separate category.

### Computing effect sizes

Each study could have a maximum of three effect sizes, one for each outcome variable. In total 63 effect sizes were computed.

For computing the effect sizes, standardised mean differences were used as effect size metric (Borenstein et al., 2009). Hedge's *g* was used as effect size unit, adjusted from Cohen's *d* to control for small sample size bias (Hedges & Olkin, 1985). In case ANCOVAs were reported, the adjusted post-test means were used to calculate the effect size. In case of a pre-post-test control group design, we intended to use the equations provided by Morris (2008) to calculate the effect size. However, none of the studies reported the correlation between pre-test and post-test results and calculating the correlation manually using formula 24 of Morris and De Shon (2002) was impossible due to missing data. Thus, sensitivity analysis would not have been possible (Borenstein et al., 2009) and this approach had to be discarded. Therefore, we used raw post-test means and standard deviations to calculate the effect sizes (Borenstein et al., 2009, p. 26; Thompson, 2007, pp. 424–425). In case neither of those data was available, information of inferential statistics (t-test, U-test) were used and the effect size was calculated with the online effect size calculator of Lipsey and Wilson (2001).

### Multiple outcomes and non-independence

Nine studies included multiple outcome measures for a single category (e.g., multiple cognitive outcomes). An example is a study that assessed scientific concept and scientific argumentation skills as outcomes (Chen, Lu & Lien, 2019). In these cases, effect sizes for each were computed and then individually pooled in a fixed meta-analysis, which were then included as pooled effect size in the main meta-analysis (Borenstein et al., 2009; Higgins et al., 2019). Additionally, 10 studies reported the data of multiple intervention or control group conditions. In those cases, a combined effect was computed, pooling the means and standard deviations using the formulas provided by Higgins et al. (2019). Two studies (Fokides, 2018a; Chang et al., 2015) reported their results for several independent groups (e.g., different grades of different schools), in which case those groups were treated as individual studies.

## **Data analysis**

For each dependent variable, one random-effects meta-analysis was conducted with the metafor package in R (Viechtbauer, 2010), using the restricted maximum-likelihood estimator for all models. The confidence interval of individual studies was compared with the confidence interval of the pulled effect and checked through sensitivity analysis to detect possible outliers (Viechtbauer & Cheung, 2010). Besides the effect size, standard deviation and its 95% confidence interval, the  $\tau^2$  for between-study variance is reported. To indicate the degree of heterogeneity of the effect sizes, Cochran's  $Q$ ,  $I^2$ , and the prediction interval were reported, which gives an overview over the effect size range (Borenstein, 2019). The  $I^2$  statistics indicates the proportion of variance between studies due to heterogeneity rather than sampling error (Borenstein, 2019).

For the moderator analysis, the mixed-effects model using the Knapp-Hartung adjustment (Knapp & Hartung, 2003) was chosen, allowing the studies within subgroups being treated as random, with subgroups being fixed and simultaneously controlling for small sample size. For categorical subgroups dummy variables were computed, whereas for continuous variables, a mixed-effects meta-regression was conducted.

## **Publication bias**

To reduce publication bias, unpublished studies were included and the publication type was added as one of the moderator variables to assess study quality. To check whether small studies with small sample sizes might have been missing, funnel plots were computed with the metafor package to check for asymmetry. Additionally, Egger's test for funnel plot asymmetry and the Duval & Tweedie's trim-and-fill procedure were conducted to adjust the effect size for publication bias (Borenstein, 2019; Van Lissa, 2019).

## **Included studies**

The 39 papers were published between 2008 and January 2020. The majority of studies came from peer-reviewed journals (90%) and all papers included information on the grade of tested participants. Regarding the subject domains in which the game intervention took place, most of them were in mathematics (40.25%), followed by biology (31%), physics (10.25%), technology (8%), chemistry (8%), and engineering (2.5%). Two-thirds used a randomised control group design, of which 15 studies randomly assigned an entire classroom to one condition, and 13 randomly assigned students within classes to a condition. Roughly one-third of included studies used a quasi-experimental design ( $k = 11$ ). Retrieving information on the background characteristics of students yielded 26 papers reporting information on gender, 4 papers reported students' migration background, 9 papers reported information on students' socioeconomic status and 9 papers included information on the amount of special educational needs students. Table B.1 in Appendix B provides additional information about the game interventions of all included studies. The most common used game design features for learning (Plass et al., 2015) in our sample were incentive systems, visual aesthetic designs, and teaching or practicing knowledge and skills.

## Additional analysis

For some moderators, there was an insufficient or relatively low number of studies to include in the moderator analyses (see 4 Results) because many studies did not report sufficient information on students' background characteristics. Therefore, we decided to complement the moderator analyses with descriptive analyses in which we qualitatively describe the potential role of the moderator variables.

## Results

### Effects and variation across studies

To answer our research question *How do effects of game-based learning differ for students with different characteristics (school level, gender, migration background, socio-economic status, special educational needs) in early STEM education?*, we conducted one meta-analysis for each outcome variable (cognition, motivation, behaviour) to examine the overall effect of game-based learning before moving into moderator analyses. The results of the random-effects meta-analyses are presented in Table 2, where the mean weighted effect size (Hedge's  $g$ ), the between-study variance, the Q-test for heterogeneity and the prediction interval are presented.

The results show, that for all three dependent variables -*cognition* ( $g = .67$ ,  $p < .001$ ), *motivation* ( $g = .51$ ,  $p < .001$ ), *behaviour* ( $g = .93$ ,  $p .01$ )- the weighted mean shows a moderate effect in favour of games compared to conventional teaching methods. This means that students who were in the intervention group scored on average 0.67 standard deviations higher in tests assessing cognitive outcomes than those who were not. The corresponding 95% confidence interval for the effect size of cognition ranges between 0.51 and 0.88, which means that the standardised mean difference in the range of comparable studies could fall anywhere in this range. The same pattern can be seen in the confidence intervals of the motivation and behaviour effect sizes, which are presented in Table 2. The Q statistic provides a test of the null hypothesis that all studies in the analysis share a common effect size (Borenstein, 2019). The Q-value is higher than the degrees of freedom for all outcome variables, with a p-value of  $<.001$ . Therefore, the null hypothesis is rejected, the true effect size is not identical in all the studies. The  $I^2$  statistic ranges from 76% to 92% that shows that the variance of observed effects are due to variance in true effects rather than sampling error.

**Table 2.** Results of the random-effects meta-analyses.

| Dependent variable | k (# students) | g (p)        | SE   | 95% CI       | $\tau^2$ (SE) | Q(p)           | df <sub>q</sub> | $I^2$ | 95% PI        |
|--------------------|----------------|--------------|------|--------------|---------------|----------------|-----------------|-------|---------------|
| Cognition          | 40 (11,363)    | 0.67 (<.001) | 0.09 | [0.51, 0.88] | 0.29 (0.08)   | 624.65 (<.001) | 39              | 92%   | [-0.38, 1.76] |
| Motivation         | 15 (9,318)     | 0.51 (<.001) | 0.09 | [0.32, 0.70] | 0.09 (0.05)   | 114.67 (<.001) | 14              | 83%   | [-0.11, 1.12] |
| Behaviour          | 5 (439)        | 0.93 (.01)   | 0.20 | [0.53, 1.32] | 0.15 (0.14)   | 18.51 (.001)   | 4               | 76%   | [0.08, 1.78]  |

k = number of studies; # students = total number of participants; g = mean weighted effect size in Hedges'  $g$ ; SE = standard error; CI = confidence interval;  $\tau^2$  = between-study variance, Q = Cochran's heterogeneity test; df = degrees of freedom Q-test;  $I^2$  = percentage of variation between studies that is due to heterogeneity rather than sampling error; PI = prediction interval

Whereas the confidence interval indicates the precision of the measured effect sizes, the prediction interval can give a better overview on the dispersion of effect sizes across populations (Borenstein, 2019). For *cognition*, the 95% prediction interval ranges from  $-0.38$  to  $+1.76$ . This means, that in the population represented by the included studies, the true effect size in 95% of cases will fall somewhere in this range. The mean effect size is moderate ( $g = 0.67$ ), indicating that the average knowledge is improved by an amount that may have substantive impact (Borenstein, 2019). However, the prediction interval shows that the dispersion of effects around this mean is substantial. Hence, there are some populations where the impact is very strong, some populations where it is moderate, and somewhere it is trivial or even negative. The same pattern can be seen in the prediction intervals for the outcome variables *motivation* ( $-0.11$ ;  $+1.12$ ) and *behaviour* ( $+0.08$ ;  $+1.78$ ).

A sensitivity analysis was performed to test for between-study heterogeneity and to check whether the observed effect size is robust and not heavily influenced by outlier studies (Viechtbauer & Cheung, 2010). The findings do not indicate any strong outlier impact on the calculated effect sizes for cognition, motivation, and behaviour.

### **Moderator analyses – background characteristics**

The results of the moderator analyses to test for differential effects of games for students with different background characteristics are presented in Tables 3 and 4 for cognition and motivation, respectively. No moderator analyses were conducted for *behaviour* because only six studies in total focused on this outcome variable, which is too few to perform a moderator analysis. All studies reported the school level and most studies reported on gender ( $k = 26$ ), but only few studies reported on the other background characteristics. For *cognition*, school level ( $k = 40$ ), gender ( $k = 23$ ), SES ( $k = 9$ ), migration background ( $k = 4$ ), SEN ( $k = 7$ ), and giftedness ( $k = 5$ ) were included as moderators. For *motivation*, moderator analyses could only be performed for school level ( $k = 15$ ), gender ( $k = 12$ ), SES ( $k = 6$ ), and SEN ( $k = 3$ ). No moderator analyses could be performed for migration background and giftedness as there were too few studies ( $k < 3$ ) reporting sufficient information.

Instead of combining the different background moderators in a multiple meta-regression, separate regression models were performed for each moderator. This approach was chosen because the focus of this meta-analysis is on the main effects of each moderator rather than their combined effects, and because combining them would mean excluding most of the studies as only very few studies included information on all moderators, thereby increasing the Type II error. To correct for Type I error, we adjusted a stricter  $p$ -value ( $p = .001$ ) for these variables. The statistics for heterogeneity ( $\tau^2$ ,  $R^2$ ,  $I^2$ ) for each regression model are reported in Tables 3 and 4.

For *cognition*, school level was found to be a significant moderator as the omnibus test of the coefficients was significant ( $p < .001$ ) and accounted for 34% of the amount of heterogeneity. The findings indicate that games are less effective for secondary school students ( $g = -0.67$ ,  $p = <.001$ ) than for primary school students. The other background variables, gender, SES, migration background, and special educational needs were not found to be significant moderators as the omnibus tests of the coefficients were not significant. More specifically, gender ( $k = 23$ ) did not have a statistically significant

moderating effect on cognition ( $p = .37$ ). Also, SES, migration background, and special educational needs were not found to be significantly moderating the effects of games on cognition. However, these results should be interpreted with caution given the small number of studies included.

For *motivation*, the school level model accounted for 22% of the heterogeneity with a significant omnibus test ( $p = .063$ ). Secondary school students seem to experience a significantly lower motivating effect than primary school students ( $g = -.37$ ,  $p = .063$ ). The omnibus tests for the study quality model and for the background characteristics were not significant, indicating that this model has no moderating effect.

### **Publication bias**

A common concern in meta-analyses is that the studies included could be a non-random subset of all studies performed. Especially since studies that report significant effects are more likely to be published than studies that do not report statistically significant results. Which in turn raises concerns that the mean effect size of any given meta-analysis could be larger than the mean effect size in all studies that were actually performed (Borenstein, 2019). One indication for such publication bias is that the effect sizes in meta-analyses are not evenly distributed around the mean and that there is a relative lack of studies with small samples and null or negative effects. To test for publication bias, the Egger regression test and Duval and Tweedie's trim-and-fill procedure were used and we added study quality characteristics in our moderator analysis.

The Egger regression test yields significant p-values for *cognition* ( $p = .004$ ), *motivation* ( $p = .002$ ), which provides evidence of a small-study effect. This could reflect the fact that the effect size is larger in smaller studies or it could reflect publication bias. The Egger regression test for *behaviour* was not significant ( $p = .284$ ). For *cognition*, the study quality characteristics model (Table 3) accounted for 36% heterogeneity with a significant omnibus test ( $p = .001$ ). Studies published in peer reviewed journals seem to have a significantly lower effect size than studies in non-peer reviewed journals ( $g = -1.216$ ,  $p = .001$ , while accounting for the other model variables). For *motivation*, this model was not significant ( $p = .855$ ).

The Duval and Tweedie's trim-and-fill procedure suggests that there may have been four studies missing for the meta-analysis on *motivation* (see, Figure 2, missing studies are indicated with white dots), and one study for the meta-analysis on *behaviour*, whereas for *cognition* no additional studies were imputed. The observed weighted mean for *motivation* is 0.51 and the adjusted weighted mean, after imputing these missing studies, is 0.43. These findings indicate that the effect would be still moderate, even if publication bias might have shifted the effect size upwards and the basic conclusion (that games have a motivating effect compared to traditional teaching) remains unchanged. A similar conclusion can be made for the meta-analysis on *behaviour*, the observed weighted mean is 0.93 and the adjusted mean, after imputing the missing study, is 0.92.

**Table 3.** Results of the moderator analysis for the dependent variable cognition.

| Moderator Variable       | k  | Estimate (SE) | p     | 95% CI         | F (df1, df2)     | p     | $\tau^2$ (SE) | I <sup>2</sup> | R <sup>2</sup> |
|--------------------------|----|---------------|-------|----------------|------------------|-------|---------------|----------------|----------------|
| School Level             |    |               |       |                | F(2, 38) = 42.61 | <.001 | 0.19 (0.06)   | 84%            | 34%            |
| Intercept                | 40 | 0.90 (0.10)   | <.001 | [0.70, 1.09]   |                  |       |               |                |                |
| Primary School (ref)     | 28 |               |       |                |                  |       |               |                |                |
| Secondary School         | 12 | -0.67 (0.17)  | <.001 | [-1.02, -0.32] |                  |       |               |                |                |
| Student Characteristics  |    |               |       |                |                  |       |               |                |                |
| Gender                   |    |               |       |                | F(1, 21) = 0.78  | .386  | 0.30 (0.11)   | 91%            | 0%             |
| Intercept                | 23 | 1.66 (1.10)   | .147  | [-0.63, 3.95]  |                  |       |               |                |                |
| Gender                   | 23 | -1.86 (2.10)  | .373  | [-6.22, 2.50]  |                  |       |               |                |                |
| Socioeconomic Status     |    |               |       |                | F(1, 7) = 0.83   | .393  | 0.44 (0.26)   | 95%            | 0%             |
| Intercept                | 9  | 0.97 (0.44)   | .068  | [-0.09, 2.02]  |                  |       |               |                |                |
| low SES                  | 9  | -0.52 (0.58)  | .393  | [-1.88, 0.84]  |                  |       |               |                |                |
| Migration background     |    |               |       |                | F(1,2) = 1.72    | .320  | 0.41 (0.44)   | 93%            | 23%            |
| Intercept                | 4  | -0.48 (0.9)   | .647  | [-4.34, 3.39]  |                  |       |               |                |                |
| Minority students        | 4  | 1.72 (1.31)   | .320  | [-3.93, 7.38]  |                  |       |               |                |                |
| Special Needs            |    |               |       |                | F(1,5) = 0.42    | .545  | 0.22 (0.16)   | 89%            | 0%             |
| Intercept                | 7  | 0.69 (0.23)   | .031  | [0.09, 1.29]   |                  |       |               |                |                |
| Special Needs students   | 7  | 0.42 (0.65)   | .545  | [-1.25, 2.09]  |                  |       |               |                |                |
| Special Needs (gifted)   |    |               |       |                | F(1,3) = 0.3     | .624  | 0.68 (0.61)   | 95%            | 0%             |
| Intercept                | 5  | 0.91 (0.46)   | .138  | [-0.53, 2.36]  |                  |       |               |                |                |
| Gifted students          | 5  | -0.53 (0.97)  | .624  | [-3.61, 2.55]  |                  |       |               |                |                |
| Quality Characteristics  |    |               |       |                | F(3, 36) = 6.73  | .001  | 0.18 (0.06)   | 85%            | 36%            |
| Intercept                | 40 | 1.79 (0.34)   | <.001 | [1.10, 2.48]   |                  |       |               |                |                |
| Study Design             |    |               |       |                |                  |       |               |                |                |
| Quasi Experimental (ref) | 15 |               |       |                |                  |       |               |                |                |
| RCT (student)            | 11 | 0.30 (0.20)   | 0.147 | [-0.11, 0.71]  |                  |       |               |                |                |
| RCT (classroom)          | 14 | -0.11 (0.18)  | 0.561 | [-0.48, 0.27]  |                  |       |               |                |                |
| Publication type         |    |               |       |                |                  |       |               |                |                |
| non peer reviewed (ref)  | 3  |               |       |                |                  |       |               |                |                |
| peer reviewed            | 37 | -1.22 (0.34)  | 0.001 | [-1.91, -0.53] |                  |       |               |                |                |

k = total number of studies; SE = standard error, CI = confidence interval, F = results of the omnibus test of all regression coefficients in a model, p = p-value of the omnibus test;  $\tau^2$  = estimated amount of residual heterogeneity; I<sup>2</sup> = percentage of unaccounted variation between studies due to heterogeneity; R<sup>2</sup> = proportion of amount of heterogeneity accounted for in the model; RCT = randomised control trial, ref = reference category

### Additional descriptive analyses

Given the limited number of studies included in the moderator analyses, we also descriptively reviewed and described the results regarding our moderators, by comparing the effect sizes for studies with different numbers of students with certain background characteristics. Thereby we aimed to see if there were some preliminary patterns emerging from the data regarding our moderators, which could not be detected in the moderator analyses. Although we cannot draw strong conclusions based on these descriptive findings, these may point to certain directions for future research.

### Socioeconomic status (SES)

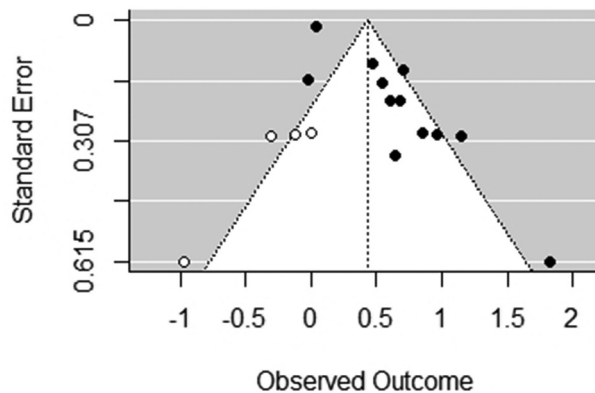
In total, only nine studies reported socioeconomic background information of their sample (see, Table 5), yet none investigated whether their game intervention had a different effect for students with different SES backgrounds. Out of those nine studies, two studies included only high SES students with the socioeconomic indicators *private school* (Yallihep & Kutlu, 2020) and *high parental education level* (Núñez Castellar, All, de Marez, & Van Looy, 2015), whereas two studies included only low SES students (Chang et al., 2015; Khan, Ahmad, & Malik, 2017) in their sample. For Chang et al. (2015), the SES



**Table 4.** Result of the moderator analysis for the dependent variable motivation.

| Moderator Variable       | k  | Estimate (SE) | p     | 95% CI        | F (df1, df2)    | p    | $\tau^2$ (SE) | $I^2$ | $R^2$ |
|--------------------------|----|---------------|-------|---------------|-----------------|------|---------------|-------|-------|
| School Level             |    |               |       |               | F(1, 13) = 4.13 | .063 | 0.07 (0.04)   | 73%   | 22%   |
| Intercept                | 15 | 0.67 (0.12)   | <.001 | [0.40, 0.94]  |                 |      |               |       |       |
| Primary School (ref)     | 9  |               |       |               |                 |      |               |       |       |
| Secondary School         | 6  | -0.37 (0.18)  | .063  | [-0.75, 0.02] |                 |      |               |       |       |
| Student Characteristics  |    |               |       |               | F(1, 10) = 2.35 | .156 | 0.08 (0.05)   | 84%   | 8%    |
| Gender                   |    |               |       |               |                 |      |               |       |       |
| Intercept                | 12 | 1.96 (0.99)   | .075  | [-0.24, 4.16] |                 |      |               |       |       |
| Gender                   | 12 | -2.96 (1.94)  | .156  | [-7.28, 1.34] |                 |      |               |       |       |
| SES                      |    |               |       |               | F(1, 4) = 1.73  | .258 | 0.04 (0.05)   | 74%   | 28%   |
| Intercept                | 6  | 0.17 (0.21)   | .464  | [-0.42, 0.77] |                 |      |               |       |       |
| low SES                  | 6  | 0.42 (0.26)   | .258  | [-0.46, 1.29] |                 |      |               |       |       |
| Special Needs            |    |               |       |               | F(1,1) = 0.33   | .666 | 0.007 (0.05)  | 23%   | 0%    |
| Intercept                | 3  | 0.46 (0.16)   | .215  | [-1.61, 2.53] |                 |      |               |       |       |
| Special Needs students   | 3  | 0.39 (0.68)   | .666  | [-8.22, 9.00] |                 |      |               |       |       |
| Quality Characteristics  |    |               |       |               | F(3, 11) = 0.26 | .855 | 0.12 (0.07)   | 77%   | 0%    |
| Intercept                | 15 | 0.65 (0.27)   | .037  | [0.05, 1.25]  |                 |      |               |       |       |
| Study Design             |    |               |       |               |                 |      |               |       |       |
| Quasi Experimental (ref) | 5  |               |       |               |                 |      |               |       |       |
| RCT (student)            | 7  | 0.11 (0.29)   | .694  | [-0.51, 0.75] |                 |      |               |       |       |
| RCT (classroom)          | 3  | -0.06 (0.34)  | .854  | [-0.82, 0.69] |                 |      |               |       |       |
| Publication type         |    |               |       |               |                 |      |               |       |       |
| non peer reviewed (ref)  | 3  |               |       |               |                 |      |               |       |       |
| peer reviewed            | 12 | -0.20 (0.34)  | .693  | [-0.96, 0.56] |                 |      |               |       |       |

k = total number of studies; SE = standard error, CI = confidence interval, F = results of the omnibus test of all regression coefficients in a model, p = p-value of the omnibus test;  $\tau^2$  = estimated amount of residual heterogeneity;  $I^2$  = percentage of unaccounted variation between studies due to heterogeneity;  $R^2$  = proportion of amount of heterogeneity accounted for in the model; RCT = randomised control trial, ref = reference category

**Figure 2.** Funnel plot after Duval & Tweedie's trim-and-fill procedure for the dependent variable motivation.

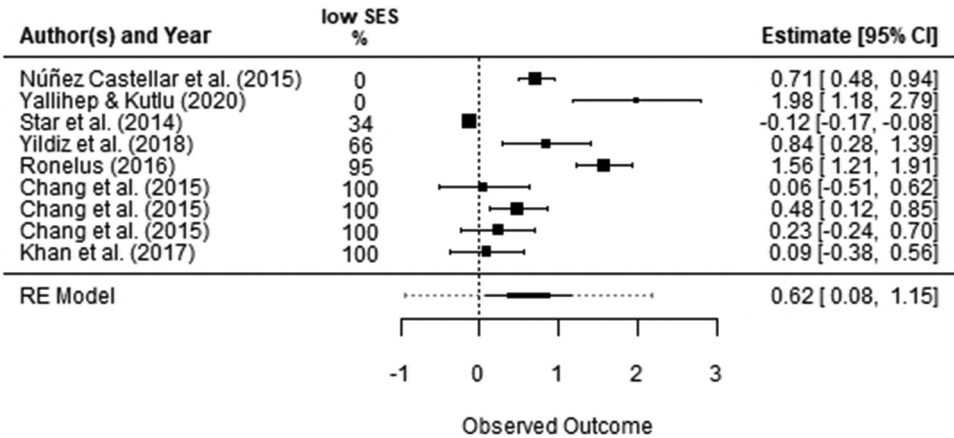
indicator *geographic level* was used, whereas the data of Khan et al. (2017) was collected in a low-cost private school in Pakistan and was coded as low SES sample through the *material and resources level*.<sup>1</sup> The remaining five studies included a heterogeneous sample, with the proportion of low SES students ranging from 34% (Star et al., 2014) to 95% (Ronelus, 2016). Among those nine studies, two investigated motivational effects only (Atwoord-Blaine, 2016; Ke, 2008), whereas the other seven studies measured either

**Table 5.** Studies reporting information on SES background.

| Author                        | Percentage low SES | Game Medium | Outcome measure                  |
|-------------------------------|--------------------|-------------|----------------------------------|
| Núñez Castellar et al. (2015) | 0%                 | computer    | cognition                        |
| Yallihep & Kutlu (2020)       | 0%                 | mobile      | cognition, motivation            |
| Star et al. (2014)            | 34%                | VR          | cognition, motivation            |
| Ke (2008)                     | 38%                | computer    | motivation                       |
| Yildiz et al. (2018)          | 66%                | paper-based | cognition                        |
| Atwood-Blaine (2018)          | 71%                | mobile      | motivation                       |
| Ronelus (2016)                | 95%                | computer    | cognition, motivation, behaviour |
| Chang et al. (2015)           | 100%               | computer    | cognition                        |
| Khan et al. (2017)            | 100%               | computer    | cognition, motivation            |

cognitive or multiple outcomes. In total, seven studies looked into cognitive effects, six studies into motivational effects, and one study measured behavioural effects of game interventions.

Even though the moderator analyses for SES failed to reach significance due to the low sample size, there seems to be a tendency of a lower effect size ( $g = -.52$ ) for the effects of games on cognition for low SES students compared to high SES students. Looking at the effect sizes of cognitive outcomes of those studies more closely, there appear to be differences in the effectiveness with the effect size ranging from  $g = -.12$  to  $g = 1.98$ , displayed in Figure 3 below. Looking particularly at the four studies that included only high or low SES students, there seems to be a difference in the effectiveness of the intervention. Whereas the two studies with high SES samples report a high effect size of  $g = 1.98$  and  $g = 0.71$  with confidence intervals in the positive range, the two studies with low SES samples report small effect sizes of  $g = 0.09$  and  $g = 0.06$  with confidence intervals covering negative as well as positive values. This observation can also be made about the data of Table 3. Hence, it may be that games are less effective for improving cognitive outcomes in low SES students, but more research is needed to test this assumption. Furthermore, the findings do not appear to demonstrate any clear pattern between SES and motivational outcomes (see forest plot in Appendix C), suggesting that games may not differ in motivational effectiveness for students with varying SES backgrounds.



**Figure 3.** Forest plot on the outcome variable cognition for the moderator variable SES.

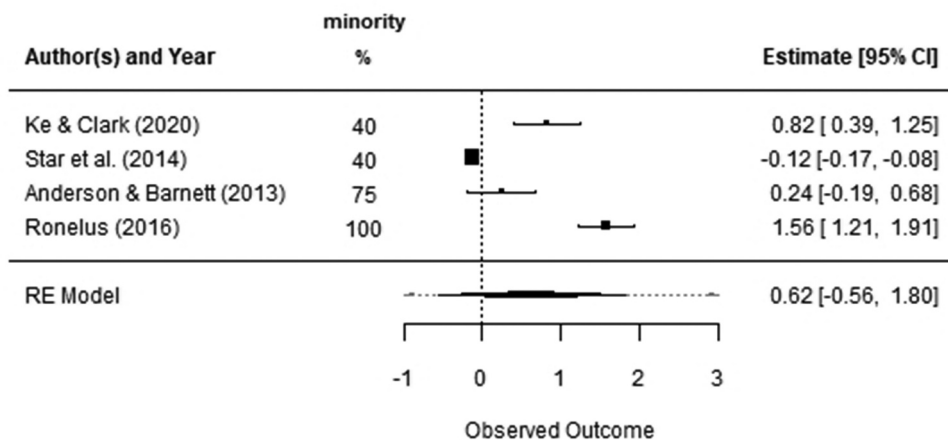
## Migration background

Four studies reported information on the minority background of students, see, Table 6. One study was conducted in a school with a student body consisting of minority students only (Ronelus, 2016), whereas for the remaining studies the proportion is ranging between 40% (Ke & Clark, 2020; Star et al., 2014) and 75% (Anderson & Barnett, 2013). All four studies investigated cognition, two additionally assessed motivation and one included behavioural outcome measures.

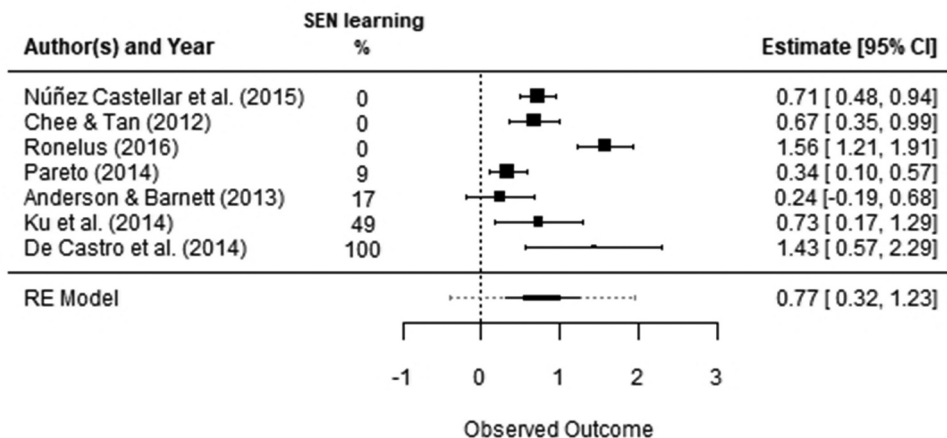
The moderator analysis was not significant ( $p = .320$ ) for the outcome variable cognition. For descriptive purposes, the effect size is displayed in the forest plot in Figure 4. The effect size for cognitive outcomes ranges from  $g = -.12$  to  $g = 1.56$ . Interestingly, the study with the highest effect size had a sample consisting of only minority students (Ronelus, 2016), whereas the lowest effect size study has the lowest sample of minority students (Star et al., 2014). For motivation, no moderator analysis could be conducted due to low number of studies. Additionally, when comparing the effect sizes in the forest plot in Appendix C, there does not seem to be any difference between the motivational outcomes for minority students in particular. In general, there does not seem to be a difference in the effectiveness of game-based learning for STEM education for this moderator.

**Table 6.** Studies reporting information on minority students.

| Author                    | Percentage migration background | Game Medium | Outcome measure                  |
|---------------------------|---------------------------------|-------------|----------------------------------|
| Ke & Clark (2020)         | 40%                             | paper-based | cognition                        |
| Star et al. (2014)        | 40%                             | VR          | cognition, motivation            |
| Anderson & Barnett (2013) | 75%                             | computer    | cognition                        |
| Ronelus (2016)            | 100%                            | computer    | cognition, motivation, behaviour |



**Figure 4.** Forest plot on the outcome variable cognition for the moderator variable migration background.



**Figure 5.** Forest plot on the outcome variable cognition for the moderator variable SEN.

### Special educational needs and giftedness

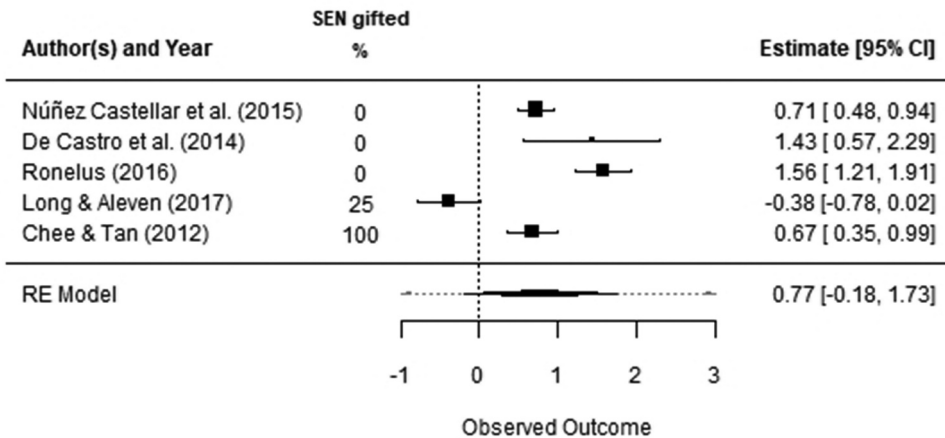
Nine studies included information on special needs status of their sample, the information can be found in Table 7. In the moderator analyses, we differentiated between SEN learning and giftedness.

Eight studies included information about the presence of students with SEN in their sample, three of which reported that there were no students with SEN, and five studies included SEN students varying from 9% to 100% of the sample. Eight studies looked into the cognitive effects, four into motivational effects and one into behavioural effects of game-based learning.

Looking specifically into the cognitive effects for SEN students, the effect sizes range from  $g = .34$  to  $g = 1.43$ , as shown in Figures 5 and 6. The highest effect size here was reached with an intervention aiming at dyscalculia students (De Castro et al., 2014) and is therefore also the only study that included a homogeneous student sample. The moderator analyses for both SEN learning and SEN gifted were not significant.

**Table 7.** Studies reporting information on SEN students.

| Author                        | Percentage SEN | Game Medium | Outcome measure                  |
|-------------------------------|----------------|-------------|----------------------------------|
| <i>SEN Learning</i>           |                |             |                                  |
| Núñez Castellar et al. (2015) | 0%             | computer    | cognition                        |
| Chee & Tan (2012)             | 0%             | computer    | cognition                        |
| Ronelus (2016)                | 0%             | computer    | cognition, motivation, behaviour |
| Pareto (2014)                 | 9%             | computer    | cognition                        |
| Anderson & Barnett (2013)     | 17%            | computer    | cognition                        |
| Ke (2008)                     | 23%            | computer    | motivation                       |
| Ku et al. (2014)              | 49%            | computer    | cognition, motivation            |
| De Castro et al. (2015)       | 100%           | computer    | cognition                        |
| <i>Gifted</i>                 |                |             |                                  |
| Núñez Castellar et al. (2015) | 0%             | computer    | cognition                        |
| De Castro et al. (2015)       | 0%             | computer    | cognition                        |
| Ronelus (2016)                | 0%             | computer    | cognition, motivation, behaviour |
| Long & Alevén (2017)          | 25%            | computer    | cognition, motivation            |
| Chee & Tan (2012)             | 100%           | computer    | cognition                        |



**Figure 6.** Forest plot on the outcome variable cognition for the moderator variable giftedness.

Of the five studies that reported on the number of gifted students, only two studies actually included gifted students, one included 25% gifted students and one study had a sample consisting only of gifted students (100%). For gifted students, the effect sizes of the two studies are  $g = -.38$  and  $g = .67$ . For these cases, it is important to point out that the sample with 100% gifted students (Chee & Tan, 2012) seemed to have a better effect on achieved learning outcomes than the sample with 25% gifted students (Long & Aleven, 2017). No difference in the effectiveness can be observed when looking at the forest plot in Figure 6. Consequently, there do not seem to be noteworthy differences of the effectiveness on cognition between special educational needs, gifted students and other students, even though some might seem to benefit slightly more from the intervention.

Finally, looking into the motivational outcomes of game-based learning, also in this case there does not seem to be any particular difference between the effectiveness for special needs and gifted students. In other words, game-based learning seems to be equally motivating for all student groups.

## Discussion

Game-based learning is increasingly implemented in primary and secondary STEM education. Even though previous meta-analyses have indicated the overall effectiveness of games for improving students' cognition, motivation, and behaviour (Lamb et al., 2018; Riopel et al., 2019; Tsai & Tsai, 2020; Vogel et al., 2006; Wouters et al., 2013), thus far little is known about whether games are equally effective for different groups of students, and thus whether some groups benefit more than others. Therefore, the aim of this meta-analysis was to gain insight whether games have different effects for students with different background characteristics (gender, age, SES, migration background, and SEN) in early STEM education. To address this aim, we first conducted three random-effects meta-analyses to compute the weighted mean effect, before moving into mixed-effect moderator analyses and complementary descriptive analyses. In short, our meta-analyses

revealed that, overall, game-based learning has positive effects on students' cognition, motivation, and behaviour, but also suggested some differences based on certain students' background characteristics, whereas others did not have any moderating effect.

## Overall findings

Aligning with findings of other meta-analyses in STEM education, which reported effect sizes ranging between 0.29 and 0.67 (Lamb et al., 2018; Riopel et al., 2019; Wouters et al., 2013), our results show that, overall, game-based learning in STEM education has beneficial effects on students' cognitive outcomes ( $g = .67$ ), their motivation ( $g = .51$ ) and behaviour ( $g = .93$ ), compared to traditional classrooms. Yet, for all three meta-analyses we found significant heterogeneity between effect sizes, suggesting that the findings substantially differed across studies. This heterogeneity was not caused by sampling error and the prediction intervals indicated substantial dispersion of effects. Thus, although the overall effect was positive, in some studies, students in the game intervention actually had more negative outcomes compared to students in traditional classrooms. This raised the question on the potential differential role of students' background characteristics.

## Moderator analyses

Unfortunately, many studies failed to report sufficient information about students' background characteristics to include them in the moderator analyses. While most studies reported students' age (school level) and the gender distribution of the sample, only few studies reported on the other background characteristics. Moreover, none of the studies included in this meta-analysis explicitly tested for differences in effects based on students' background. This complicated and limited the possibilities for performing moderator analyses. This was especially the case for the outcome category *behaviour*. The total number of included studies for behaviour was too low to perform any moderator analyses for this outcome variable. For *cognition* and *motivation*, only for the moderators age and gender, a sufficient number of studies was included to be able to draw solid conclusions. Given the small number of studies reporting on students' SES, migration background, or SEN, the findings of these moderator analyses need to be considered as preliminary findings and interpreted with caution. Therefore, we combined these moderator analyses with a more descriptive review of the findings to provide some first indications regarding potential directions of moderating effects. Below, we discuss the findings for each moderator.

**Age.** The moderator analyses for students' school level (as an indication of their age) included a sufficient number of studies to be able to draw conclusions. Our results indicate that games are less beneficial for secondary school students compared to primary school students, with regard to cognition ( $g = -.67$ ) and motivation ( $g = -.37$ ) for STEM. In other words, especially primary school students seem to benefit from the use of STEM games compared to traditional classroom settings. These results contrast previous findings that indicated that game-based learning is equally effective for all age groups (Tokac et al., 2018; Tsai & Tsai, 2020) or less beneficial for younger students (Riopel et al., 2019; Vogel et al., 2006; Wouters et al., 2013). None of the previous reviews did specifically compare primary with secondary school students and in most cases, primary and

secondary school students were combined and compared with adults in higher education. However, since previous research indicated that the motivation for STEM subjects decreases already from upper elementary school (Renninger et al., 2015; Shin et al., 2019), it may be that comparing primary with secondary school students yields a differential effect that could possibly be explained with the general decrease of interest which is common in the population of secondary school students. Hence, these findings indicate that interventions using games to enhance cognition and motivation may be especially effective when implemented early.

*Gender.* Gender on the other hand had no significant moderating effect on cognition ( $p = .37$ ) and motivation ( $p = .16$ ), meaning there is no recordable difference in the effectiveness of game-based learning for boys and girls. Those results align with the result of Vogel et al. (2006), showing no significant difference in the effectiveness of game-based learning. This means, even though boys and girls have different preferences for game types (Homer et al., 2012) and play them in different frequencies (Hygen et al., 2019), they equally benefit from game-based learning interventions in STEM subjects.

*SES.* Similar to what Scherer and Siddiq (2019) found when examining the effect of ICT use, students with low socioeconomic backgrounds may learn less when game-based learning is used compared to a traditional classroom setting. Even though the moderator analysis for SES failed to reach significance, there seems to be a tendency of a lower effect size ( $g = -.52$ ) for the effects of games on cognition for low SES students compared to high SES students that could also be observed in the descriptive analysis. For motivation on the other hand, no difference in the effectiveness of game-based learning was found, indicating that games can be equally motivating for low and high SES students. One possible explanation could be the lack of home resources of low SES students, which would mean that they are less accustomed to the use of technology in general. The game-based learning intervention could be perceived as fun and exciting, raising the motivation for the subject, but due to the lack of ICT skills compared to high SES students, cognitive overload could interfere with successfully learning from these interventions. This aligns with the results of Hanghøj et al. (2018), where at-risk students were motivated through game-based learning in school, but did not learn more through the intervention.

*Migration background.* Previous research on ICT literacy indicated that minority students report lower ICT skills compared to majority students (Volman et al., 2005). Therefore, we expected to find a moderating effect for game-based learning interventions. The moderator analyses, however, were not significant and no differences in the effectiveness were found in the descriptive analysis. Further research is needed to investigate possible differential effects for this moderator, since we could only include four studies in this meta-analysis that reported information on minority background students.

*SEN and Giftedness.* The moderator analyses for both SEN learning and SEN gifted were not significant and no noteworthy differences in effectiveness were found through the descriptive analysis. Games have been particularly designed and used as treatment for special educational needs students and have been shown to be effective (Ke & Moon, 2018; Lu et al., 2018; Ronimus et al., 2019). But none of these studies looked into the effects of game-based learning intervention compared to traditional teaching raising the question whether similar results could also be seen in these interventions. On the other hand, we suspected that gifted students might not be challenged enough in



game-based learning interventions leading to under-stimulation that could negatively influence the observed outcomes. Our results are preliminary, but it seems that game-based learning does not increase pre-existing inequalities and both SEN and gifted students benefit from game-based learning interventions as much as students without SEN.

These results have several practical implications. First, games are a suitable tool to motivate all students for a STEM subject and therefore might be a great way to reach also disengaged students in a classroom. However, since not every student seems to be able to learn well with game-based learning, teachers have to be careful about how to design their lessons, taking into consideration that some students might need additional support to be able to benefit from using games. Furthermore, future research could investigate whether additional instruction or the use of dynamic difficulty adjustment (Van Oostendorp et al., 2014) within a game can counteract the disadvantages that low SES students seem to be facing. Although the number of included studies in our meta-analysis was rather small, our results show that it is important to take the different characteristics (particularly age and SES) of students into consideration when planning a game-based learning intervention to avoid contributing to pre-existing differences.

## Limitations

The majority of the included studies did not report sufficient demographic information of their sample, apart from the information reported for age and gender. Therefore, some of the moderator analyses could not be conducted or were performed with low numbers of studies. One approach to deal with such missing data would be to manually insert estimates (Borenstein et al., 2009), but this had to be discarded in our case as the number of missing studies was too high. Another approach would be to use a random forests algorithm for meta-analyses to explore heterogeneity and to identify relevant moderators (Van Lissa, 2020). Again, this approach was not feasible for our sample as the number of missing data was simply too high to draw any conclusions. Therefore, the results of the moderator analyses on the other background characteristics and the additional descriptive analyses can only be seen as a first indication of potential differential effects or the lack of thereof.

One general difficulty of conducting a meta-analysis is to find all available studies that fit the inclusion criteria. With searching in several multidisciplinary databases we tried to find all papers available, but we cannot fully exclude the possibility that the choice of databases has limited the amount of studies that were found. Moreover, our exclusion criteria might have limited the number of available studies. One example is the exclusion of non-English articles which might have excluded relevant papers in other languages. Another example is the exclusion of articles that investigated a different age group. It could be of interest to investigate the possible differential effects of game-based learning in other age groups and settings as well, as games that are being used in e.g., employee training might show a similar differential effect depending on the background of the employee.

We adopted a broad understanding of games and did not look into a media-comparison approach to distinguish between e.g., paper-based games or virtual game environments, which might yield different effects. Additionally, we did not look into specific game features that could have an effect on the outcome variables and might moderate the effects, as this was beyond the scope of the present meta-analysis.

## Implications for theory, research and practice

The findings of this meta-analysis suggest several implications for theory, future research, and practice.

*Implications for theory.* Game-based learning is a multidisciplinary field and as such, research on game-based learning is very diverse in terms of the underlying theories. Many studies on game-based learning adopt theories from other fields, for example, theories on multimedia learning, motivation, and, to a lesser extent theories on instructional design. However, theories on inclusive education, for example, the universal design for learning approach (UDL; CAST, 2018) have, to our knowledge, not been incorporated in studies on games. These theories can provide insights into how to make games suitable for diverse student groups. Incorporating such theories in intervention studies and game design could help to make games optimally suitable for different student groups and maximise the beneficial effects.

*Implications for future research.* In this meta-analysis, it was not possible to investigate possible interaction effects of students' background characteristics with certain game features due to our limited sample size. Nevertheless, it would be relevant to gain more insight about the effectiveness of game-based learning for different student groups. To allow future meta-analyses to look into these types of interactions, future empirical studies need to be more elaborate on their sample characteristics. Moreover, future meta-analyses could investigate the differential effect of games not only in STEM education but also include other domains. This could lead to a higher sample size and could allow additional moderator analyses to investigate possible interaction effects between certain game mechanics and player characteristics.

In addition, empirical studies are needed which explicitly compare the effects of games between student groups. Even more, games can strongly differ from one another and they may also elicit different effects in different disciplines. There may be interactions between game features, disciplines, and student characteristics (e.g., certain games may be particularly effective for certain groups in a certain domain). Empirical studies are needed to address such interactions, to provide more insights into which types of games, in which disciplines, work best for which students. Moreover, future research could also examine the factors that may underlie different effectiveness for different groups, such as poor self-regulation skills or differences in experiences with games. Gaining insight into the exact causes of differences can help to overcome these differences in effectiveness by adjusting the game design or game instructions.

*Implications for practice.* Overall, our findings confirmed findings of previous meta-analyses (Lamb et al., 2018; Riopel et al., 2019; Vogel et al., 2006; Wouters et al., 2013; Zainuddin et al., 2020) and suggested that games can be a helpful tool to improve students' learning outcomes, motivation, and behaviour. However, our findings also warrant some caution. That is, despite the inconclusiveness of our findings, the findings

do seem to suggest that the effects of games can differ between student groups. This means that educators need to be aware that games that have been shown to be effective in general, may not be as effective for their specific population or the effects may vary between students. Hence, educators need to carefully consider if and which games they may implement for their student population and they need to monitor continuously if the games yield the desired effects, and if there are students who do not benefit from the game in terms of their cognition, motivation, or behaviour. In these cases, additional instruction (i.e., pretraining) or an alternative method may be required.

More specifically, the findings suggested that the games may be less effective for low SES students than for their higher SES counterparts. Hence, implementing games in diverse classrooms may exacerbate existing gaps between students with different socio-economic backgrounds and needs to be considered carefully. Moreover, our findings show that game-based learning in STEM education is effective in general, but especially younger students seem to benefit from the use of games. This means that games should be implemented early on to be most effective, particularly to improve students' motivation for STEM subjects which might help to intervene with the general decrease of interest in these subjects later on.

## Conclusion

The present study quantitatively summarised the results of 39 studies about the effects of games on STEM learning, motivation, and behaviour compared to conventional classroom settings in K1-8. In general, it can be concluded that the students in the game interventions achieve significantly higher learning outcomes, report more motivation, and behavioural change compared to students in traditional classrooms. Particularly students in primary education (K1-6) seem to benefit from game interventions in STEM education. The high heterogeneity between effect sizes in our sample, however, indicates that the implementation of games as additional teaching tools should be planned with caution, as not every student is equally benefiting from game-based learning. Our study provides some preliminary support for the assumption that low socioeconomic background students might learn less with game-based learning than students with a higher socioeconomic status, whereas gender, migration background, and special educational needs did not seem to have any differential effect. More empirical research is needed that reports and investigates the effectiveness of game-based learning for heterogeneous student groups. Future meta-analyses will then be able to provide more substantiated results.

## Note

1. Private schools in Pakistan are divided between low-cost and elite schools and make up two thirds of all available middle schools (Asian Development Bank, 2019). Even though those schools are labelled private, one quarter is missing books (UNESCO, 2007). Compared to public schools they still have better resources such as running water and electricity, but following our definition of low SES the low-cost private school was coded as such.

## Acknowledgments

The authors would like to warmly thank Selma van Aken for her help in the data collection.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the European Union's EU Framework Programme for Research and Innovation Horizon 2020 under Grant Agreement No. 812716 [Horizon 2020 Framework Programme 812716].

## Notes on contributors

### *Biographical Note*

**Michaela Arztmann** is a doctoral researcher focusing on how immersive learning environments can be used to interest children with diverse backgrounds for chemistry.

**Lisette Hornstra** works as an associate professor at the education department of Utrecht University. Her research interests include students' motivation for school, social and ethnic diversity in schools, and teaching high-ability students.

**Johan Jeuring** is head of department and professor of Software Technology for Learning and Teaching at the Institute of Information and Computing Sciences of Utrecht University, and professor of Software Technology at the School of Computer Science of the Open University. His research focusses on software technology for learning and teaching.

**Liesbeth Kester** is full professor Educational Sciences at Utrecht University (UU). She is chair of the UU division Education. Recently she obtained research grants from the European Commission (Marie-Sklodowska-Curie Innovative Training Network within the Horizon 2020 Programme) for research into game-based learning and the Netherlands Initiative for Education Research for research into learning analytics. Her expertise includes instructional design, multimedia learning, hypermedia learning, personalised learning, cognitive aspects of learning, including for example, testing and retention, and designing and developing flexible learning environments.

## ORCID

Michaela Arztmann  <http://orcid.org/0000-0002-4453-7892>

Liesbeth Kester  <http://orcid.org/0000-0003-0482-0391>

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author, Arztmann, M., upon request.

## References

- Aesaert, K., Van Nijlen, D., Vanderlinde, R., Tondeur, J., Devlieger, I., & van Braak, J. (2015). The contribution of pupil, classroom and school level characteristics to primary school pupils' ICT competences: A performance-based approach. *Computers & Education*, 87, 55–69. <https://doi.org/10.1016/j.compedu.2015.03.014>
- American Psychological Association, Task Force on Socioeconomic Status. (2007). *Report of the APA task force on socioeconomic status*.
- Anderson, J. L., & Barnett, M. (2013). Learning physics with digital game simulations in middle school science. *Journal of Science Education and Technology*, 22(6), 914–926. <https://doi.org/10.1007/s10956-013-9438-8>
- Asian Development Bank (2019). *School education in Pakistan. A sector assessment*. Asian Development Bank. <https://dx.doi.org/10.22617/TCS190039>
- Atwood-Blaine, D. (2016). *The effect of playing a science centre-based mobile game: Affective outcomes and gender differences* (Doctoral dissertation). <http://hdl.handle.net/1808/21680>
- Ayman R., Sharaf N., Ahmed G., & Abdennadher S. (2018). MiniColon; Teaching kids computational thinking using an interactive serious game. In S. Göbel et al. (eds.) *Serious games. JCSG 2018. Lecture notes in computer science* (pp 79–90). Springer, Cham. [https://doi.org/10.1007/978-3-030-02762-9\\_9](https://doi.org/10.1007/978-3-030-02762-9_9)
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295x.84.2.191>
- Bayir, E., & Evmez, S. (2019). The effects of inquiry-based experiment-integrated science games among secondary school students. *International Journal of Evaluation and Research in Education (IJERE)*, 8(3), 434. <https://doi.org/10.11591/ijere.v8i3.20244>
- Betancur, L., Votruba-Drzal, E., & Schunn, C. (2018). Socioeconomic gaps in science achievement. *International Journal of STEM Education*, 5(1). <https://doi.org/10.1186/s40594-018-0132-5>
- Betancur, L., Votruba-Drzal, E., & Schunn, C. (2018). Socioeconomic gaps in science achievement. *International Journal of STEM Education*, 5(1), 38. <https://doi.org/10.1186/s40594-018-0132-5>
- Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them*. Biostat, Inc.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470743386>
- CAST (2018). *Universal design for learning guidelines version 2.2*. <https://udlguidelines.cast.org/>
- Chang, M., Evans, M. A., Kim, S., Norton, A., & Samur, Y. (2015). Differential effects of learning games on mathematics proficiency. *Educational Media International*, 52(1), 47–57. <https://doi.org/10.1080/09523987.2015.1005427>
- Chee, Y.S., & Tan, K. (2012). Becoming chemists through game-based inquiry learning: the case of 'Legends of Alkhimia'. *Electronic Journal of e-Learning*, 10, 185–198.
- Chen, C.-H., Liu, G.-Z., & Hwang, G.-J. (2016). Interaction between gaming and multistage guiding strategies on students' field trip mobile learning performance and motivation. *British Journal of Educational Technology*, 47, 1032–1050. <https://doi.org/10.1111/bjet.12270>
- Chen, Y.-C., Lu, Y.-L., & Lien, C.-J. (2019). Learning environments with different levels of technological engagement: A comparison of game-based, video-based, and traditional instruction on students' learning. *Interactive Learning Environments*, 1–17. <https://doi.org/10.1080/10494820.2019.1628781>
- Chen, M.-H., Tsai, S.-T., & Chang, C.-C. (2019). Effects of game-based instruction on the results of primary school children taking a natural science course. *Education Sciences*, 9(2), 79. <https://doi.org/10.3390/educsci9020079>
- Cheng, M.-T., Su, T., Huang, W.-Y. and Chen, J.-H. (2014). An educational game for learning human immunology: What do students learn and how do they perceive?. *British Journal of Educational Technology*, 45, 820–833. <https://doi.org/10.1111/bjet.12098>
- Chu, H.-C., & Chang, S.-C. (2013). Developing an educational computer game for migratory bird identification based on a two-tier test approach. *Educational Technology Research and Development*, 62(2), 147–161. <https://doi.org/10.1007/s11423-013-9323-4>

- De Castro, M. V., Bissaco, M. A. S., Panccioni, B. M., Rodrigues, S. C. M., & Domingues, A. M. (2014). Effect of a virtual environment on the development of mathematical skills in children with dyscalculia. *PLoS ONE*, 9(7), e103354. <https://doi.org/10.1371/journal.pone.0103354>
- Demirel, T., & Karakus Yilmaz, T. (2018). The effects of mind games in maths and grammar courses on the achievements and perceived problem-solving skills of secondary school students. *British Journal of Educational Technology*, 50(3), 1482–1494. <https://doi.org/10.1111/bjet.12624>
- Dieser, O., & Bogner, F. X. (2015). Young people's cognitive achievement as fostered by hands-on-centred environmental education. *Environmental Education Research*, 22(7), 943–957. <https://doi.org/10.1080/13504622.2015.1054265>
- European Agency for Special Needs and Inclusive Education (2019). *Preventing school failure: A review of the literature*. [https://www.european-agency.org/sites/default/files/psf\\_literature\\_review\\_v2.pdf](https://www.european-agency.org/sites/default/files/psf_literature_review_v2.pdf)
- European Commission (2015). *Directorate for research and innovation (2015). Science education for responsible citizenship*. Publications Office of the European Union. <https://doi.org/10.2777/12626>
- European Commission (2021). *Digital skills and jobs*. <https://digital-strategy.ec.europa.eu/en/policies/digital-skills-and-jobs>
- Fine-Davis, M., & Faas, D. (2014). Equality and diversity in the classroom: A comparison of students' and teachers' attitudes in six European countries. *Social Indicators Research*, 119(3), 1319–1334. <https://doi.org/10.1007/s11205-013-0547-9>
- Fiorella, L., Kuhlmann, S., & Vogel-Walcutt, J. J. (2018). Effects of playing an educational maths game that incorporates learning by teaching. *Journal of Educational Computing Research*, 57(6), 1495–1512. <https://doi.org/10.1177/0735633118797133>
- Fokides, E. (2018a). Digital educational games and mathematics. Results of a case study in primary school settings. *Education and Information Technologies*, 23(2), 851–867. <https://doi.org/10.1007/s10639-017-9639-5>
- Fokides, E. (2018b). Teaching basic programming concepts to young primary school students using tablets. *International Journal of Mobile and Blended Learning*, 10(1), 34–47. <https://doi.org/10.4018/ijmbli.2018010103>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Garneli, V., Giannakos, M., & Chorianopoulos, K. (2016). Serious games as a malleable learning medium: The effects of narrative, gameplay, and making on students' performance and attitudes. *British Journal of Educational Technology*, 48(3), 842–859. <https://doi.org/10.1111/bjet.12455>
- Garon-Carrier, G., Boivin, M., Lemelin, J.-P., Kovas, Y., Parent, S., Séguin, J. R., Vitaro, F., Tremblay, R. E., & Dionne, G. (2018). Early developmental trajectories of number knowledge and math achievement from 4 to 10 years: Low-persistent profile and early-life predictors. *Journal of School Psychology*, 68, 84–98. <https://doi.org/10.1016/j.jsp.2018.02.004>
- Gnambs, T. (2021). The development of gender differences in information and communication technology (ICT) literacy in middle adolescence. *Computers in Human Behavior*, 114, 106533. <https://doi.org/10.1016/j.chb.2020.106533>
- Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, 102, 202–223. <https://doi.org/10.1016/j.compedu.2016.09.001>
- Hanghøj, T., Lieberoth, A., & Misfeld, M. (2018). Can cooperative video games encourage social and motivational inclusion of at-risk students? *British Journal of Educational Technology*, 49(4), 775–799. <https://doi.org/10.1111/bjet.12642>
- Harrington, B., & O'Connell, M. (2016). video games as virtual teachers: Prosocial video game use by children and adolescents from different socioeconomic groups is associated with increased empathy and prosocial behaviour. *Computers in Human Behavior*, 63, 650–658. <https://doi.org/10.1016/j.chb.2016.05.062>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.



- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127. [https://doi.org/10.1207/s15326985ep4102\\_4](https://doi.org/10.1207/s15326985ep4102_4)
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., et al. (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). The Cochrane Collaboration.
- Homer, B. D., Hayward, E. O., Frye, J., & Plass, J. L. (2012). Gender and player characteristics in video game play or preadolescents. *Computers in Human Behavior*, 28(5), 1782–1789. <https://doi.org/10.1016/j.chb.2012.04.018>
- Homer, B. D., Raffaele, C., & Henderson, H. (2020). Games as playful learning: Implications of developmental theory for game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 25–52). The MIT Press.
- Hwang, G.-J., & Chen, C.-H. (2016). Influences of an inquiry-based ubiquitous gaming design on students' learning achievements, motivation, behavioural patterns, and tendency towards critical thinking and problem solving. *British Journal of Educational Technology*, 48(4), 950–971. <https://doi.org/10.1111/bjet.12464>
- Hwang, G.-J., Sung, H.-Y., Hung, C.-M., Yang, L.-H. & Huang, I. (2013), A knowledge engineering approach to developing educational computer games for improving students' differentiating knowledge. *British Journal of Educational Technology*, 44, 183–196. <https://doi.org/10.1111/j.1467-8535.2012.01285.x>
- Hygen, B. W., Belsky, J., Stenseng, F., Skalicka, V., Kvande, M. N., Zahl-Thanem, T., & Wichstrøm, L. (2019). Time spent gaming and social competence in children: Reciprocal effects across childhood. *Child Development*, 91(3), 861–875. <https://doi.org/10.1111/cdev.13243>
- Ke, F. (2008). Computer games application within alternative classroom goal structures: Cognitive, metacognitive, and affective evaluation. *Educational Technology Research and Development*, 56(5–6), 539–556. <https://doi.org/10.1007/s11423-008-9086-5>
- Ke, F., M. Clark, K. (2020). Game-based multimodal representations and mathematical problem solving. *International Journal of Science and Mathematics Education*, 18, 103–122. <https://doi.org/10.1007/s10763-018-9938-3>
- Ke, F., & Moon, J. (2018). Virtual collaborative gaming as social skills training for high-functioning autistic children. *British Journal of Educational Technology*, 49(4), 728–741. <https://doi.org/10.1111/bjet.12626>
- Khan, A., Ahmad, F., & Malik, M.M. (2017). Use of digital game based learning and gamification in secondary school science: The effect on student engagement, learning and gender difference. *Education and Information Technologies*, 22, 2767–2804. <https://doi.org/10.1007/s10639-017-9622-1>
- Kinzie, M. B., & Joseph, D. R. D. (2008). *Gender differences in game activity preferences of middle school children: Implications for educational game design*. Educational Technology Research and Development. <https://doi.org/10.1007/s11423-007-9076-z>
- Klopfer, E., & Thompson, M. (2020). Game-based learning in science, technology, engineering, and mathematics. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 387–408). The MIT Press.
- Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17), 2693–2710. <https://doi.org/10.1002/sim.1482>
- Ku, O., Chen, S.Y., Wu, D.H., Lao, A.C.C., & Chan, T.-W. (2014). The effects of game-based learning on mathematical confidence and performance: high ability vs. low ability. *Journal of Educational Technology & Society*, 17(3), 65–78. <http://www.jstor.org/stable/jeductechsoci.17.3.65>
- Lamb, R. L., Annetta, L., Firestone, J., & Etio, E. (2018). A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. *Computers in Human Behavior*, 80, 158–167. <https://doi.org/10.1016/j.chb.2017.10.040>
- Lee, J. Y., Donkers, J., Jarodzka, H., & van Merriënboer, J. J. G. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior*, 99, 268–277. <https://doi.org/10.1016/j.chb.2019.05.035>
- Li, H., Liu, J., Zhang, D., & Liu, H. (2021). Examining the relationships between cognitive activation, self-efficacy, socioeconomic status, and achievement in mathematics: A multi-level analysis. *British Journal of Educational Psychology*, 91(1), 101–126. <https://doi.org/10.1111/bjep.12351>



- Lie, R., Selcen Guzey, S., & Moore, T. J. (2019). Implementing Engineering in Diverse Upper Elementary and Middle School Science Classrooms: Student Learning and Attitudes. *Journal of Science Education and Technology*, 28(2), 104–117. <https://doi.org/10.1007/s10956-018-9751-3>
- Lin, C.-H., Liu, E. Z.-F., Chen, Y.-L., Liou, P.-Y., Chang, M., Wu, C.-H., & Yuan, S.-M. (2013). Game-based remedial instruction in mastery learning for upper-primary school students. *Journal of Educational Technology & Society*, 16(2), 271–281. <http://www.jstor.org/stable/jeductechsoci.16.2.271>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.
- Loderer, K., Pekrun, R., & Plass, J. L. (2020). Emotional foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 111–151). The MIT Press.
- Long, Y., & Aleven, V. (2017). Educational game and intelligent tutoring system: A classroom study and comparative design analysis. *ACM Transactions on Computer-Human Interaction*. <https://doi.org/10.1145/3057889>
- Lu, A., Chan, S., Cai, Y., Huang, L., Nay, Z. T., & Goei, S. L. (2018). Learning through VR gaming with virtual pink dolphins for children with ASD. *Interactive Learning Environments*, 26(6), 718–728. <https://doi.org/10.1080/10494820.2017.1399149>
- Marriott, L. K., Coppola, L. A., Mitchell, S. H., Bouwma-Gearhart, J. L., Chen, Z., et al. (2019). Opposing effects of impulsivity and mindset on sources of self-efficacy and STEM interest among adolescents. *PLoS One*, 14(8). <https://doi.org/10.1371/journal.pone.0201939>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386. <https://doi.org/10.1177/1094428106291059>
- Morris, S. B., & De Shon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-group designs. *Psychological Methods*, 7(1), 105–125. <https://doi.org/10.1037/1082-989x.7.1.105>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International results in mathematics and science*. TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/international-results/>
- Nietfeld, J. L. (2020). Predicting transfer from a game-based learning environment. *Computers & Education*, 146, 103780. <https://doi.org/10.1016/j.compedu.2019.103780>
- Núñez Castellar, E., All, A., de Marez, L., & Van Looy, J. (2015). Cognitive abilities, digital games and arithmetic performance enhancement: A study comparing the effects of a maths game and paper exercises. *Computers & Education*, 85, 123–133. <https://doi.org/10.1016/j.compedu.2014.12.021>
- OECD. (2019). *PISA 2018 results (volume II): Where all students can succeed*. <https://doi.org/10.1787/b5fd1b8f-en>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan – A web and mobile app for systematic reviews. *Systematic Reviews*, 5(1). <https://doi.org/10.1186/s13643-016-0384-4>
- Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *International Journal of Artificial Intelligence in Education*, 24(3), 251–283. <https://doi.org/10.1007/s40593-014-0018-8>
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, 50(4), 258–283. <https://doi.org/10.1080/00461520.2015.1122533>
- Plass, J. L., Homer, B. D., Mayer, R. E., & Kinzer, C. K. (2020). Theoretical foundations of game-based and playful learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 3–24). The MIT Press.
- Renninger, K. A., Costello Kensey, C. N., Stevens, S. J., & Lehman, D. L. (2015). Perceptions of science and their role in the development of interest. In K. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest in mathematics and science learning* (pp. 93–110). American Educational Research Association. [www.jstor.org/stable/j.ctt1s474j0](http://www.jstor.org/stable/j.ctt1s474j0)

- Riopel, M., Nenciovici, L., Potvin, P., Chastenay, P., Charland, P., et al. (2019). Impact of serious games on science learning achievement compared with more conventional instruction: An overview and a meta-analysis. *Studies in Science Education*, 55(2), 169–214. <https://doi.org/10.1080/03057267.2019.1722420>
- Ronelus, W.J. (2016). *Using multi-user virtual simulation to promote science content: Mastery, scientific reasoning, and academic self-efficacy in fifth grade science* (Doctoral dissertation). Available from ProQuest Dissertations and Theses. PhDT168R
- Ronimus, M., Eklund, K., Pesu, L., & Lyytinen, H. (2019). Supporting struggling readers with digital game-based learning. *Educational Technology Research and Development*, 67(3), 639–663. <https://doi.org/10.1007/s11423-019-09658-3>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066x.55.1.68>
- Ryan, R. M., & Rigby, C. S. (2020). Motivational foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 153–176). The MIT Press.
- Scherer, R., & Siddiq, F. (2019). The relation between students' socioeconomic status and ICT literacy: Findings from a meta-analysis. *Computers & Education*, 138, 13–32. <https://doi.org/10.1016/j.compedu.2019.04.011>
- Schunk, D. H., & DiBenedetto, M. K. (2020). Social cognitive theory, self-efficacy, and students with disabilities. implications for students with learning disabilities, reading disabilities, and attention-deficit/hyperactivity disorder. In A. J. Martin, R. A. Sperling, & K. J. Newton (Eds.), *Handbook of educational psychology and students with special needs* (pp. 243–261). Routledge.
- Shi, A., Wang, Y., & Ding, N. (2019). The effect of game-based immersive virtual reality learning environment on learning outcomes: Designing an intrinsic integrated educational game for pre-class learning. *Interactive Learning Environments*, 1–14. <https://doi.org/10.1080/10494820.2019.1681467>
- 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>
- Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., Ventura, M., & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education*, 86, 224–235. <https://doi.org/10.1016/j.compedu.2015.08.001>
- Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, 42(1), 70–83. <https://doi.org/10.1037/0012-1649.42.1.70>
- Star, J.R., Chen, J.A., Taylor, M.W., et al. (2014). Studying technology-based strategies for enhancing motivation in mathematics. *International Journal of STEM Education*, 1(7). <https://doi.org/10.1186/2196-7822-1-7>
- Stewart, J., Bleumers, L., Van Looy, J., Mariln, I., All, A., et al. (2013). *The potential of digital games for empowerment and social inclusion of groups at risk of social and economic exclusion: Evidence and opportunity for policy*. Publications Office of the European Union. <https://doi.org/10.2791/88148>
- Su, C.-H., & Cheng, C.-H. (2014). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning*, 31(3), 268–286. <https://doi.org/10.1111/jcal.12088>
- Sung, H.-Y., Hwang, G.-J., Wu, P.-H., & Lin, D.-Q. (2018). Facilitating deep-strategy behaviours and positive learning performances in science inquiry activities with a 3D experiential gaming approach. *Interactive Learning Environments*, 26(8), 1053–1073. <https://doi.org/10.1080/10494820.2018.1437049>
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, 44(5), 423–432. <https://doi.org/10.1002/pits.20234>

- Tokac, U., Novak, E., & Thompson, C. G. (2018). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. <https://doi.org/10.1111/jcal.12347>
- Tsai, C-Y, Lin, H-s, & Liu, S-C. (2020). The effect of pedagogical GAME model on students' PISA scientific competencies. *Journal of Computer Assisted Learning*, 36, 359–369 . <https://doi.org/10.1111/jcal.12406>
- Tsai, Y., & Tsai, C. (2020). A meta-analysis of research on digital game-based science learning. *Journal of Computer Assisted Learning*, 36(3), 280–294. <https://doi.org/10.1111/jcal.12430>
- Turan, G. Y., Köklükaya, A. N., & Yıldırım, E. G. (2019). Improving matter and heat subjects learning through genuine designed educational games. *International Journal of Science and Mathematics Education*, 18(1), 19–42. <https://doi.org/10.1007/s10763-018-09945-0>
- Turner, S. L., Jeong, J. R., Sims, M. D., Dade, S. N., & Froman Reid, M. (2019). SES, gender, and STEM career interests, goals, and actions: A test of SCCT. *Journal of Career Assessment*, 27(1), 134–150. <https://doi.org/10.1177/1069072717748665>
- UNESCO. (2007) . *The education system in Pakistan*.
- Van Lissa, C. J. (2019). *Doing meta-analysis in R and exploring heterogeneity using metaforest*. <https://cjvanlissa.github.io/Doing-Meta-Analysis-in-R/>
- Van Lissa, C. J. (2020). Small sample meta-analyses: exploring heterogeneity using metaForest. In R. Van De Schoot & M. Miočević (Eds.), *Small sample size solutions (open access): A guide for applied researchers and practitioners*. CRC Press (pp.186–202). <https://www.crcpress.com/Small-Sample-Size-Solutions-Open-Access-A-Guide-for-Applied-Researchers/Schoot-Miocevic/p/book/9780367222222>
- Van Oostendorp, H., Van der Spek, E. D., & Linssen, J. (2014). Adapting the complexity level of a serious game to the proficiency of players. *EAI Endorsed Transactions on Serious Games*, 1(2), 1–8. <https://doi.org/10.4108/sq.1.2.e5>
- Verhaeghe, J. P., Vanlaar, G., Knipprath, H., De Fraine, B., & Van Damme, J. (2018). Can group composition effects explain socioeconomic and ethic achievement gaps in primary education? *Studies in Educational Evaluation*, 57, 6–15. <https://doi.org/10.1016/j.stueduc.2017.07.006>
- 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.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229–243. <https://doi.org/10.2190/flhv-k4wa-wpvq-h0ym>
- Volman, M., Van Eck, E., Heemskerk, J., & Kuiper, E. (2005). New technologies, new differences. Gender and ethnic differences in pupils' use of ICT in primary and secondary education. *Computers & Education*, 45(1), 35–55. [https://doi.org/10.1016/s0360-1315\(04\)00072-7](https://doi.org/10.1016/s0360-1315(04)00072-7)
- Wang, S.-Y., Chang, S.-C., Hwang, G.-J., & Chen, P.-Y. (2017). A microworld-based role-playing game development approach to engaging students in interactive, enjoyable, and effective mathematics learning. *Interactive Learning Environments*, 26(3), 411–423. <https://doi.org/10.1080/10494820.2017.1337038>
- Watson, J. (2018). *An experimental study on the effects of a gamified software intervention in mathematics achievement among sixth grade students* (Doctoral dissertation). [https://nsuworks.nova.edu/fse\\_etd/222](https://nsuworks.nova.edu/fse_etd/222)
- Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249–265. <https://doi.org/10.2196/18644>
- Yallihp, M., Kutlu, B. (2020) Mobile serious games: Effects on students' understanding of programming concepts and attitudes towards information technology. *Education and Information Technologies*, 25, 1237–1254. <https://doi.org/10.1007/s10639-019-10008-2>

Yildiz, E., Şimsek, Ü., & Agdas, H. (2018). The effects of educational game-integrated group research method on academic achievement, attitude towards school, and retention of knowledge in teaching regulatory system. *Journal of Turkish Science Education*, 15(3), 91–105.

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, 30. <https://doi.org/10.1016/j.edurev.2020.100326>

References included in the Meta-Analysis

## Appendix A: Search string sample

The following text shows one example of a search string that was used in Ovid (PsychINFO). The first paragraph shows the search syntax on the term game-based learning with all relevant synonyms which is used in the subject headings search. The second paragraph contains a similar search syntax, but now the terms are for the free text search. '1 or 2' is combining both paragraphs, meaning that one of those terms should be found either in the free text of the paper or in the subject heading. This approach is then followed through with all other relevant terms for the meta-analysis.

exp Digital Game-Based Learning/ or exp Computer Games/ or exp Computer Assisted Instruction/ or exp Virtual Reality/ or exp Augmented Reality/ or exp Simulation Games/ or virtual classrooms/

(game\* or game-based learning or serious game\* or educational game\* or video game\* or computer game\* or simulation game\* or augmented reality or virtual reality or computer assisted instruction or digital game-based learning or virtual classrooms or AR or VR or XR or MR).ti,ab,id.

1 or 2

exp Motivation/ or exp Achievement/ or Interests/ or exp Self-Efficacy/ or exp Values/ or exp Goals/ or Involvement/ or exp Achievement Motivation/ or exp Performance/ or exp Participation/

(motivation or interest or self-efficacy or value\* or goal\* or involvement\* or achievement\* or performance\* or participation).ti,ab,id.

4 or 5

3 and 6

exp science education/ or exp STEM/ or exp chemistry/ or exp engineering/ or exp physics/

(science education or STEM or chemistry or engineering or physics).ti,ab,id.

8 or 9

7 and 10

## Appendix B

**Table 1.:** Additional study information.

| Authors            | Game medium | Included game design elements for learning   | Number of interventions | Outcome variable        | g              |
|--------------------|-------------|--|-------------------------|-------------------------|----------------|
| Anderson & Barnett | computer    | Visual aesthetic design (virtual world, in-world objects)<br>Teach new knowledge and skills (laws of electromagnetism)   | multiple                | Cognition               | 0.242          |
| Atwood-Blaine      | mobile      | Incentive system (points, badges)<br>Narrative (solving a mystery)<br>Teach new knowledge and skills (science and engineering practices)<br>Develop 21 <sup>st</sup> century skills (collaboration, problem solving)                                   | single                  | Motivation              | 0.424          |
| Ayman et al.       | computer    | Visual aesthetic design (virtual world, in-world character and objects)<br>Narrative (trapped on an Island and needs to escape)<br>Teach new knowledge and skills (programming)  | single                  | Cognition<br>Motivation | 1.894<br>1.818 |
| Bayir & Emez       | paper       | Teach new knowledge and skills (light, sound, and electricity)   | multiple                | Cognition               | 1.599          |
| Chang et al.       | computer    | Visual aesthetic design (in-world objects)<br>Practice and reinforce knowledge and skills (fractions)  | multiple                | Cognition               | 0.233          |
| Chee & Tan         | computer    | Visual aesthetic design (virtual worlds, in-world characters and objects)<br>Narrative (trapped in an underground chemistry lab)<br>Teach new knowledge and skills (chemistry separation)  | multiple                | Cognition               | 0.671          |
| Chen, Liu & Hwang  | tablet      | Incentive system (scores)<br>Visual aesthetic design (virtual world)<br>Practice and reinforce knowledge and skills (plant recognition)  | single                  | Cognition<br>Motivation | 0.745<br>0.217 |
| Chen, Tsai & Chang | computer    | Visual aesthetic design (avatar, virtual world)<br>Teach new knowledge and skills (carbon footprint)   | multiple                | Cognition               | 0.203          |
| Chen, Lu & Lien    | computer    | Visual aesthetic design (avatar, virtual world, in-world objects)<br>Teach new knowledge and skills (light)  | multiple                | Cognition               | 0.886          |
| Cheng et al.       | computer    | Incentive system (scores)<br>Visual aesthetic design (in-world character and objects)<br>Narrative (alien attack)<br>Teach new knowledge and skills (human immune response)  | multiple                | Cognition               | 0.217          |
| Chu & Chang        | computer    | Visual aesthetic design (virtual world, in-world character and objects)<br>Narrative (treasure hunt)<br>Teach new knowledge and skills (migratory bird identification)   | multiple                | Cognition<br>Motivation | 0.806<br>0.967 |
| De Castro et al.   | computer    | Incentive system (score)<br>Visual aesthetic design (avatar, in-world character and objects)<br>Narrative (rescue turtle Tom)<br>Practice and reinforce knowledge and skills (number identification, even and odd numbers, four arithmetic operations) | multiple                | Cognition               | 1.427          |

(Continued)

**Table 1.:** (Continued).

| Authors                            | Game medium | Included game design elements for learning  | Number of interventions | Outcome variable                     | g                       |
|------------------------------------|-------------|---|-------------------------|--------------------------------------|-------------------------|
| Demirel & Yilmaz                   | paper       | Incentive system (points)<br>Practice and reinforce knowledge and skills (algebra, geometry)  | multiple                | Cognition                            | 0.891                   |
| Dieser & Bogner                    | paper       | Narrative (imitate hunting behaviour of wolves)<br>Teach new knowledge and skills (forest ecology and species)  | single                  | Cognition                            | 1.426                   |
| Fiorella, Kuhlmann & Vogel-Walcutt | computer    | Visual aesthetic design (virtual world, in-game character)<br>Narrative (space adventure)<br>Practice and reinforce knowledge and skills (equations, volume, mean)  | multiple                | Cognition<br>Motivation              | 0.281<br>–<br>0.017     |
| Fokides (2018a)                    | computer    | Practice and reinforce knowledge and skills (weight, decimals, money)   | multiple                | Cognition                            | 0.700                   |
| Fokides (2018b)                    | tablet      | Incentive system (coins)<br>Visual aesthetic design (virtual world, in-world objects)<br>Teach new knowledge and skills (programming)   | multiple                | Cognition                            | 1.328                   |
| Garneli et al.                     | computer    | Incentive system (diamonds)<br>Visual aesthetic design (in-world character and objects)<br>Narrative (dog has been kidnapped)<br>Practice and reinforce knowledge and skills (addition and subtraction of integers) | single                  | Motivation<br>Behaviour              | 0.060<br>0.560          |
| Hwang & Chen                       | tablet      | Incentive system (scores, ranking)<br>Visual aesthetic design (in-world character and objects)<br>Learn new knowledge and skills (environmental pollution)  | multiple                | Cognition<br>Motivation<br>Behaviour | 0.695<br>0.611<br>0.417 |
| Hwang et al.                       | computer    | Visual aesthetic design (in-world character and objects)<br>Narrative (king is looking for a cure to save his kingdom)<br>Learn new knowledge and skills (plant properties)   | single                  | Cognition<br>Behaviour               | 0.586<br>1.335          |
| Ke                                 | computer    | Visual aesthetic design (in-world character and objects)<br>Practice and reinforce knowledge and skills (equations, measurements, coordinates)  | multiple                | Motivation                           | 0.466                   |
| Ke & Clark                         | computer    | Visual aesthetic design (in-world objects)<br>Narrative (rebuild a disaster-damaged space)<br>Practice and reinforce knowledge and skills (ratio, volume, algebra)  | multiple                | Cognition                            | 0.819                   |
| Khan, Ahmad & Malik                | computer    | Incentive system (points)<br>Visual aesthetic design (in-game character and objects)<br>Musical score (audio)<br>Learn new knowledge and skills (reaction)  | multiple                | Cognition<br>Motivation              | 0.092<br>0.703          |
| Ku et al.                          | computer    | Incentive system (scores)<br>Visual aesthetic design (in-game character)<br>Practice and reinforce knowledge and skills (mental calculation)  | multiple                | Cognition<br>Motivation              | 0.732<br>0.849          |
| Lin et al.                         | computer    | Virtual aesthetic design (virtual world)<br>Practice and reinforce knowledge and skills (circle)  | single                  | Cognition                            | 0.453                   |
| Long & Aleven                      | computer    | Virtual aesthetic design (in-world character and objects)<br>Learn new knowledge and skills (algebra)   | multiple                | Cognition<br>Motivation              | –0.379<br>0.679         |

(Continued)

**Table 1.:** (Continued).

| Authors                     | Game medium | Included game design elements for learning  | Number of interventions | Outcome variable                     | g                       |
|-----------------------------|-------------|---|-------------------------|--------------------------------------|-------------------------|
| Núñez Castellar et al.      | computer    | Incentive system (score)<br>Virtual aesthetic design (in-world character and objects)<br>Practice and reinforce knowledge and skills (mental arithmetic skills)   | multiple                | Cognition                            | 0.710                   |
| Pareto                      | computer    | Incentive system (points)<br>Practice and reinforce knowledge and skills (algebra, strategic thinking)  | multiple                | Cognition                            | 0.337                   |
| Ronelus                     | computer    | Visual aesthetic design (virtual world, in-world objects)<br>Learn new knowledge and skills (ecosystem and habitat)   | multiple                | Cognition<br>Motivation<br>Behaviour | 1.564<br>0.543<br>1.120 |
| Shi, Wang & Ding            | VR          | Incentive system (scores)<br>Visual aesthetic design (virtual world, in-world characters and objects)<br>Practice and reinforce knowledge and skills (quadratic function)   | single                  | Cognition                            | −0.494                  |
| Star et al.                 | VR          | Visual aesthetic design (virtual world, in-world objects)<br>Narrative (space rescue mission)<br>Practice and reinforce knowledge and skills (algebraic reasoning)  | multiple                | Cognition<br>Motivation              | −0.123<br>0.035         |
| Su & Cheng                  | mobile      | Visual aesthetic design (virtual world)<br>Practice and reinforce knowledge and skills (insects)  | multiple                | Cognition                            | 0.812                   |
| Sung, Hwang & Lin           | computer    | Visual aesthetic design (virtual world, in-world-objects)<br>Musical score (audio)<br>Practice and reinforce knowledge and skills (rocks and minerals)<br>Develop 21 <sup>st</sup> century skills (problem solving) | multiple                | Cognition<br>Motivation<br>Behaviour | 0.589<br>1.148<br>1.363 |
| Tsai, Lin & Liu             | computer    | Incentive system (points, rewards)<br>Learn new knowledge and skills (ocean acidification, scientific thinking)   | multiple                | Cognition                            | 0.727                   |
| Turan, Köklükaya & Yildirim | paper       | Incentive system (scores)<br>Practice and reinforce knowledge and skills (heat, matter, and fuels)  | multiple                | Cognition                            | 0.771                   |
| Wang et al.                 | computer    | Visual aesthetic design (in-world character and objects)<br>Narrative (Tortoise and the Hare)<br>Practice and reinforce knowledge and skills (speed)  | single                  | Cognition                            | 0.212                   |
| Watson                      | computer    | Incentive system (points, medals)<br>Practice and reinforce knowledge and skills (numbers, measurements, geometry, statistics, algebra, probability)  | multiple                | Cognition                            | 2.501                   |
| Yallihep & Kutlu            | mobile      | Learn new knowledge and skills (programming)  | multiple                | Cognition<br>Motivation              | 1.980<br>0.637          |
| Yildiz, Simsek & Agdas      | paper       | Incentive system (scores)<br>Practice and reinforce knowledge and skills (endocrine glands)   | multiple                | Cognition                            | 0.837                   |



Appendix C: Forest plots for the moderator analyses on motivation

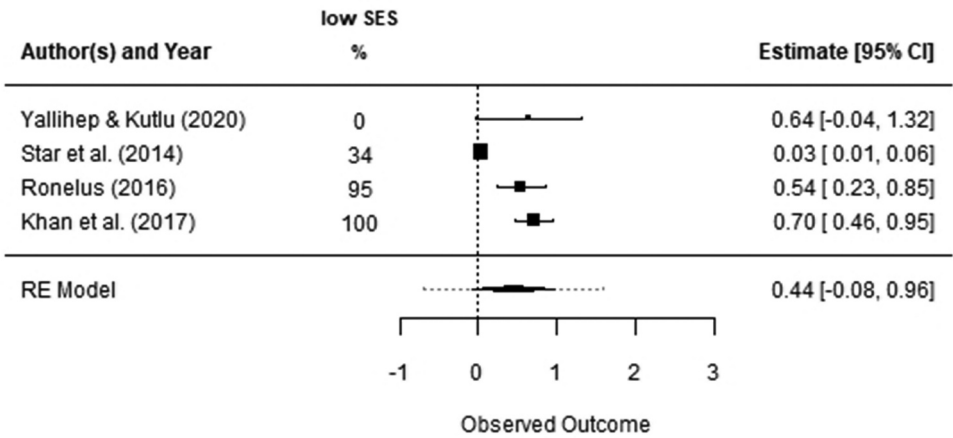


Figure C 1: Forest plot on the outcome variable motivation for the moderator variable SES.

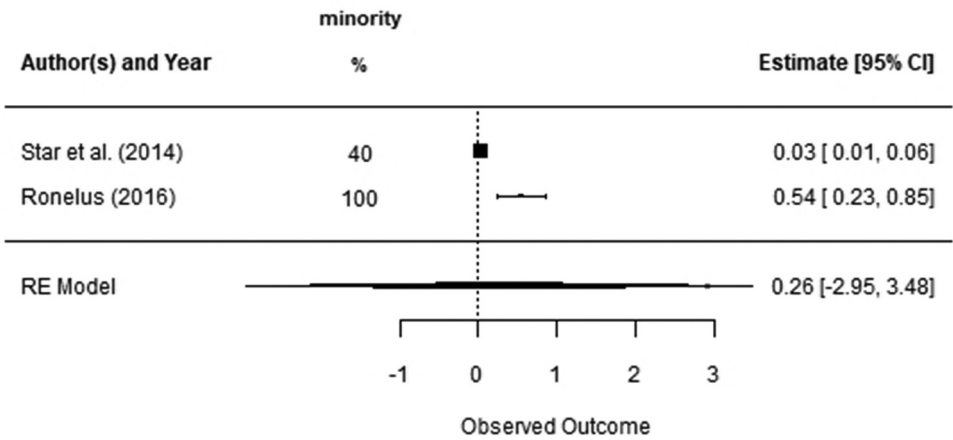
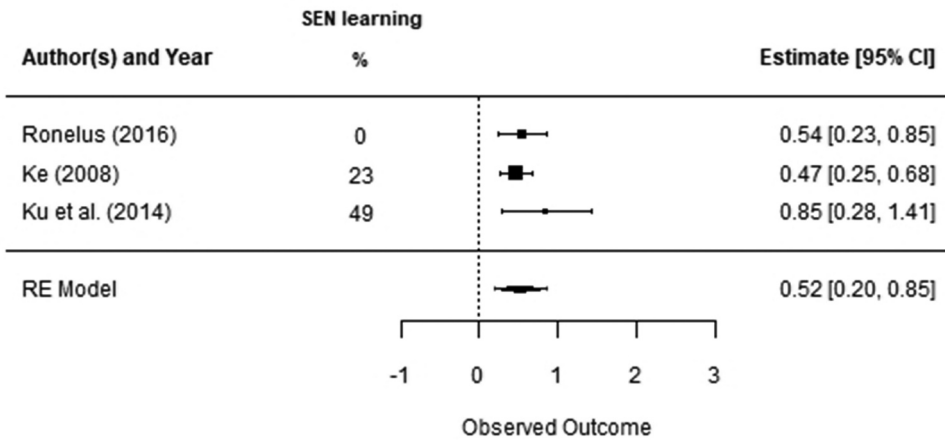
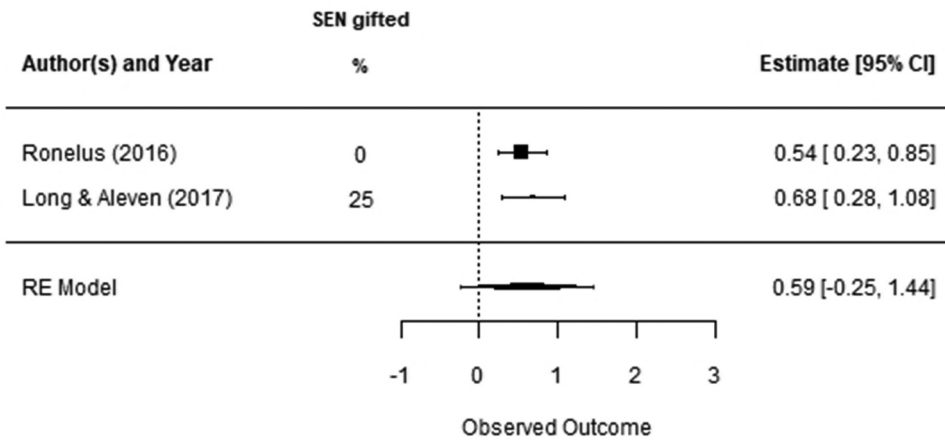


Figure C 2: Forest plot on the outcome variable motivation for the moderator variable migration background.



**Figure C. 3:** Forest plot on the outcome variable motivation for the moderator variable SEN.



**Figure C 4:** Forest plot on the outcome variable motivation for the moderator variable giftedness.