European Journal of Science and Mathematics Education, 2026, **14**(4), 525-546
ISSN: 2301-251X (Online)

Review Article

Effects of science education on computational thinking: A meta-analysis

Gülbin Kıyıcı [1*]
0000-0002-5402-0117

Havva Yamak [2]
0000-0002-1666-0746

[1] Department of Mathematics and Science Education, Faculty of Education, Manisa Celal Bayar University, Manisa, TURKEY
[2] Department of Mathematics and Science Education, Gazi Faculty of Education, Gazi University, Ankara, TURKEY
[*] Corresponding author: gulbin.kiyici@cbu.edu.tr

Citation: Kıyıcı, G., & Yamak, H. (2026). Effects of science education on computational thinking: A meta-analysis. *European Journal of Science and Mathematics Education, 14*(4), 525-546. https://doi.org/10.30935/scimath/18904

ARTICLE INFO

Received: 23 Dec 2025

Accepted: 17 Jun 2026

ABSTRACT

This meta-analysis provides a comprehensive overview of the findings of studies in science education conducted between 2006 and 2021 (April) to improve computational thinking. The research process began with a literature review and the establishment of eligibility criteria. Following this, a coding form was developed to ensure the reliability of the coding, a pilot coding process was conducted, and the coding form and coding guide were finalized. This process concluded with data analysis, evaluation of findings, and reporting (investigator coding and parallel coding). The overall mean effect size (d = 0.714) for the 32 primary studies that met the inclusion criteria resulted in a moderate effect. Additionally, many moderator variables were determined (publication type, language, country, the status of the pilot study, year of publication, method, design, model, sampling method of the study group, the demographic structure of the study group, school type of study group, application research area, duration of application, the person performing the application, using the computer, coding, robotics, algorithm and flipped classroom method in the application, type of measurement tool, and the person who developed the measurement tools). Finally, an in-depth discussion of how these variables identified as moderators relate to their effectiveness was included.

Keywords: computational thinking, meta-analysis, science education

INTRODUCTION

Information processing and problem-solving processes in the human brain are generally referred to as computational thinking (Wing, 2006). Computational thinking involves establishing systematic structures for solving complex problems, analyzing these structures, and developing solutions. This process considers engineering and scientific thinking approaches, along with concepts such as computability, intelligence, reasoning, and human behavior. In this respect, computational thinking largely overlaps with mathematical thinking (Korkmaz et al., 2015; Wing, 2006). In other studies in the literature, computational thinking is a fundamental skill that everyone should have to cope with problems, regardless of the field of study (Hsu et al., 2018; Voogt et al., 2015; Wing, 2006). Therefore, many different definitions explain the concept of computational thinking from different perspectives. When these definitions are examined, computational thinking is expressed in a way that includes many skills and/or features.

Hsu et al. (2018) classified computational thinking according to the thinking process as in **Figure 1**. These thinking processes are abstraction, problem-solving, algorithm design, modeling, automation, efficiency and performance, data analysis, debugging and error, detection, data collection, visualization, representative data, connection to other fields, decomposition, conditional, logic, parallelization, transformation, pattern generalization, simulation. **Figure 1** shows that the computational thinking process consists of 18 elements.

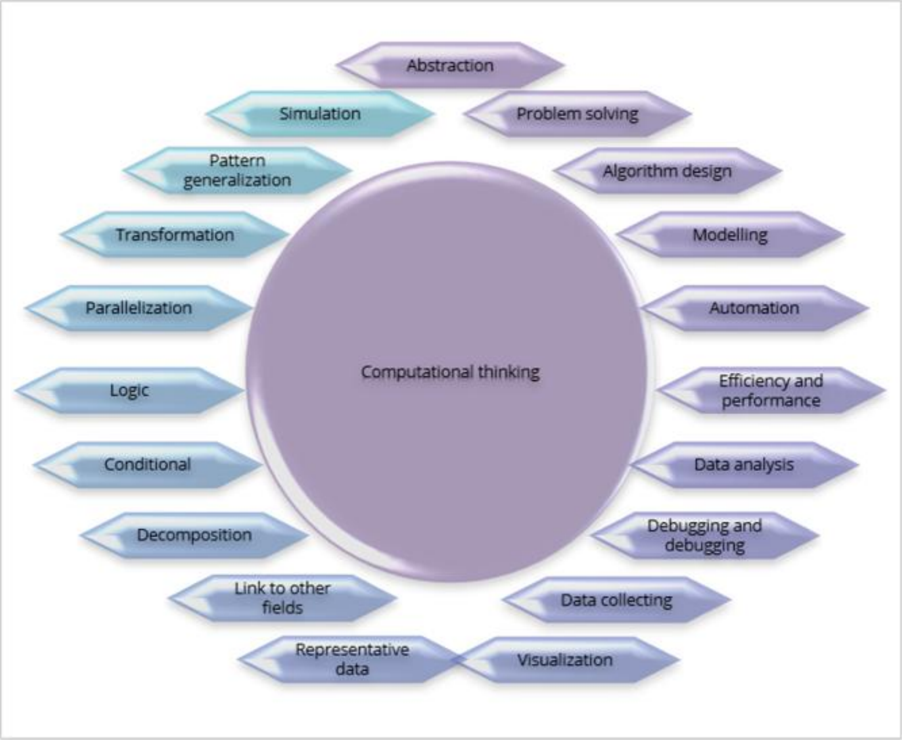


Figure 1. Computational thinking process (Hsu et al., 2018)

Computational thinking contributes to the development of basic skills such as reading, writing, and arithmetic, as well as thinking skills such as problem-solving, analysis, and decision-making (Wing, 2006).

Computational thinking skills enable individuals to be successful not only in numerical calculations or computer-related situations but also in different areas. A study suggests that even without elective courses in science and mathematics education, students can learn computational thinking skills related to planned thinking (Rohaeti & Huda, 2025).

The basic procedures/algorithms and abstraction elements of computational thinking are effective in helping students reach higher quality products in their learning processes and develop thinking skills that will help them comprehend different academic subjects such as mathematics, science, and language (Grover & Pea, 2017). Abstraction appears in different forms in every field. In mathematics, students expressing the quantities of objects with numbers and applying arithmetic algorithms are examples of abstraction. In science, examples can be given such as matching the name of a compound with its formula (e.g., glucose: $C_6H_{12}O_6$), applying chemical algorithms (e.g., writing the products of a reaction given the reactants), and creating systems (e.g., adding reactions in aerobic respiration). Therefore, individuals with a high level of computational thinking can show high academic success in different fields (Lei et al., 2020).

In addition to problem-solving, critical thinking, and scientific process skills in science education, findings from studies on computational thinking in recent years make the integration of computational thinking into this field an important study topic. For example, in a study conducted for this purpose, robotics and coding-based STEM activities in students contributed to the development of students' critical thinking, computational thinking, creative thinking, and problem-solving skills (Akyol, 2020).

Science education is a field that prepares individuals for life and ensures that they are prepared for all the situations they will encounter throughout life (Cömert & Kıyıcı, 2013; Yavru & Gürdal, 1998). Therefore, learning science gives individuals the necessary skills to meet their natural needs more easily, adapt to the environment, and cope with nature. Researchers highlighted that the concept of computational thinking is best possible with science education, which provides learning in the context of real-world problems, and there is a relationship between science learning and computational thinking (Guzdial 1994; Hambrusch et al. 2009; Perkins & Simmons, 1988). In a study showing that components of computational thinking can deepen the learning of science content, the dynamic nature of computational models allows students to both “observe” and “interact” with target events in real-time, and also perform conceptual learning (Akşit, 2018).

In recent years, science education studies in which computational thinking applications have been made provide important information about how this integration can be done (Guzdial, 1994; Hambrusch et al., 2009; Jona et al., 2014; Luo et al., 2020; NRC, 2000; Pintrich et al., 1991; Redish & Wilson, 1993). As a result, it is important to examine the relationship between science education and computational thinking skills. For this purpose, in this study, first of all, a comprehensive overview of the findings of the studies in the field of science education aimed at improving computational thinking was made and these studies were analyzed by meta-analysis with a holistic approach. Of these studies, 33 primary studies were identified that met the meta-analysis inclusion criteria. A coding form, where these studies are methodologically reliable, their findings are reported objectively, their researchers do not act in a biased way, the measurement tools and the method followed are suitable for the research, and they are designed following the experimental design rules. The information obtained from the survey is limited to the information reported in primary studies. Then, after transforming the experimental results of published empirical studies into comparable effect sizes, the average effect of applications in science education on computational thinking was examined. In addition, the effect of the determined moderator variables on computational thinking in science education was determined.

In this study, answers to the following questions were sought.

What are the effect sizes when primary studies in which science education applications are conducted to improve the computational thinking levels of learners (robotics, coding, algorithms, computer, and flipped classroom applications) are examined by meta-analysis method?

According to the results of primary research, which distinguishing characteristics influence science education?

METHODS

A meta-analysis study was conducted in this study. Meta-analysis is a statistical method used to combine experimental findings obtained from individual studies on similar subjects, to synthesize and interpret them in the form of effect size (Card, 2012; Littell et al., 2008).

Evaluating heterogeneity in meta-analysis is an important issue because whether there is real heterogeneity (variability between studies), which statistical model will be applied in the meta-analysis is decided (Borenstein et al., 2009). Testing heterogeneity, which is one of the main purposes of meta-analysis, also shows the presence of moderator variables, and some quantitative criteria are used to determine whether they exist (Göktaş, 2017). These are Cochran Q, Higgins I^2 and T^2 (tau squared).

In meta-analysis studies, it is necessary to determine the full inclusion/exclusion criteria that will enable us to answer the research question (Card, 2012). After reviewing the relevant literature, the compliance of the studies obtained with the inclusion and exclusion criteria is evaluated. The coding form, which is accepted as the data collection tool of the meta-analysis study, is created (Üstün, 2012). Judgment, which is inevitable when coding the coder’s primary work, can reveal differences in how elements on the coding page are encoded. This difference can occur in two ways: encoder reliability and consistency between different encoders (Lipsey & Wilson, 2001). Intra-coder reliability is calculated by examining the consistency of scores given by the same individual. It is mostly determined by the average compliance rate and Cohen’s kappa coefficient (Vevea & Zelinsky, 2019). Intercoder reliability (ICR) is the degree of agreement or consistency between multiple coders (Cohen & Swerdlik, 2018).

The validity of the results of a meta-analysis study may be low if the studies included in the meta-analysis are biased, no matter how impeccable other methodological issues are (Becker, 2005). In this context, publication bias can occur when the sample research conducted according to the literature does not represent the universe of systematically completed studies.

In this meta-analysis study, a rigorous research process was followed to collect, analyze and summarize the empirical evidence relevant to the research questions. Results are reported according to “preferred reporting items for systematic reviews and meta-analyses”, also known as “PRISMA” (Liberati et al., 2009). The

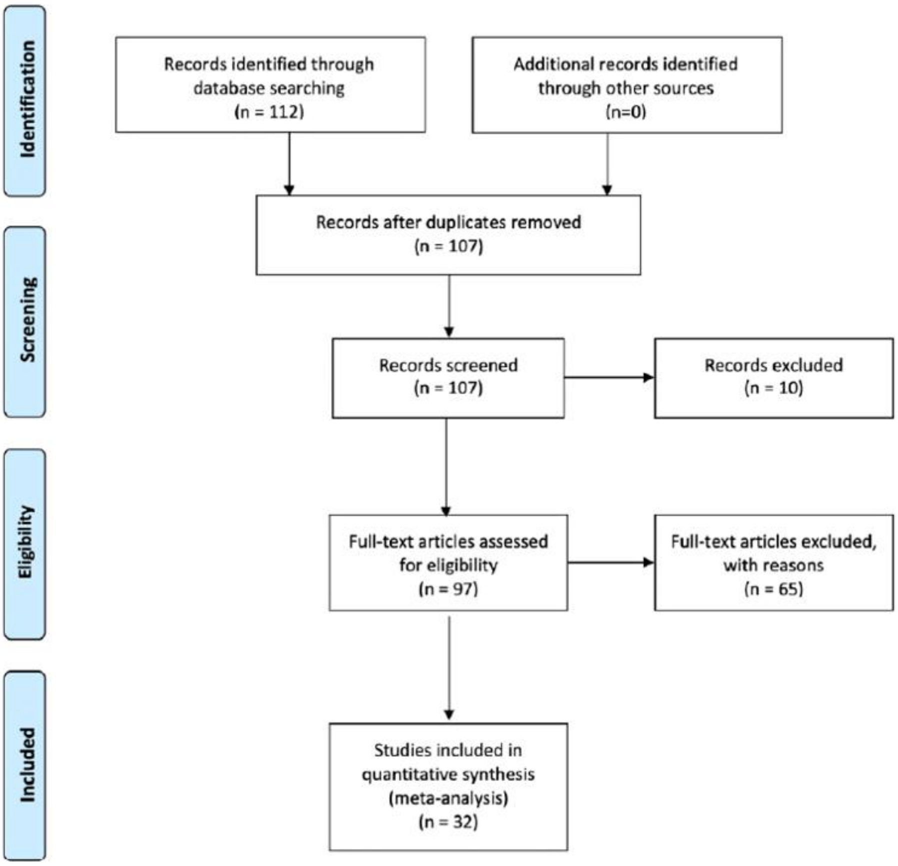


Figure 2. PRISMA 2009 flowchart for meta-analysis (Page et al., 2021)

recommendations of the Cochrane handbook for systematic reviews of primary research were followed (Higgins et al., 2003)

Literature Review

The following stages were applied in the literature review process of this research:

1. Databases were determined for the literature review of the meta-analysis. The thesis catalog of the Higher Education Institution, ULAKBİM, pages of scientific journals, ProQuest Digital Dissertations, Proquest EBSCOhost, ERIC, ScienceDirect, Web of Science, Taylor & Francis Online, Scopus, and JSTOR databases were used.
2. Research was conducted based on the key concepts of "Science Education", "Science Teaching", "Science Learning", "STEM" and "Computational Thinking" and relevant literature was determined.
3. These works of literature have been categorized according to the eligibility criteria given below. In a meta-analysis, it is necessary to determine the full inclusion/exclusion criteria that will enable us to answer our research question in line with the purpose of the research (Card, 2012).

PRISMA Flowchart

Figure 2 shows the 4 stages of the meta-analysis process and the number of primary studies remaining at each stage.

Inclusion/Exclusion Criteria

Only studies with the following characteristics were included in the meta-analysis:

1. Containing applications or variables related to science education,
2. Studies included in the meta-analysis were published between 2006 and 2020 (because Jeannette Wing's article "computational thinking" in 2006 was the pioneering publication).

Table 1. Cohen's kappa statistics for intra-coder and intercoder reliability.

	κ	Asymptotic standard error	Approximate Tc	Significance level
In encoder				
Cohen's kappa	1.000	0.000	28.203	0.000
Valid cases	330			
Studies in Turkish				
Cohen's kappa	0.779	0.046	13.714	0.000
Valid cases	132			
Studies in English				
Cohen's kappa	0.688	0.440	16.470	0.000
Valid cases	1,199			

Note. Not assuming the null hypothesis and using the asymptotic standard error assuming the null hypothesis

3. Computational thinking level being the dependent variable,
4. Published or unpublished master's and doctoral theses, articles in academic (peer-reviewed) journals, papers presented at conferences and symposiums,
5. Experimental and quasi-experimental research model,
6. Reporting effect size or numerical data necessary to calculate effect size,
7. Studies to be included in the analysis must be published in Turkish or English.

Research studies with the following characteristics were excluded from the scope of meta-analysis:

1. Articles published in journals other than peer-reviewed academic journals,
2. Theses that report the same findings and have been published as both theses and articles,
3. Studies that do not include statistical data or raw scores that would allow for the calculation of effect size,
4. Studies that do not meet the inclusion criteria were excluded from the scope of the research.

Data Coding and Encoder Reliability

Characteristics expressing the features of the research that were decided to be included in the analysis were coded in a systematically collected coding form by transforming them into quantitative data. Coding was carried out by two lecturers who were experts in the field of coding reliability.

In this study, the percentage of agreement and Cohen's kappa statistics were used for the ICR analysis, assuming that the coded items, the encodings of the coders, and the categories of the items were independent of each other. The average within-encoder agreement rate was 1.0, and the mean agreement rate between the researcher and coder_Tr was 0.856. Additionally, the average agreement rate between the researcher and coder_Eng was calculated as 0.801. Looking at **Table 1**, Cohen's kappa value for the two encodings (within the encoder) made by the researcher was 1.00 and it was statistically significant ($p < 0.05$).

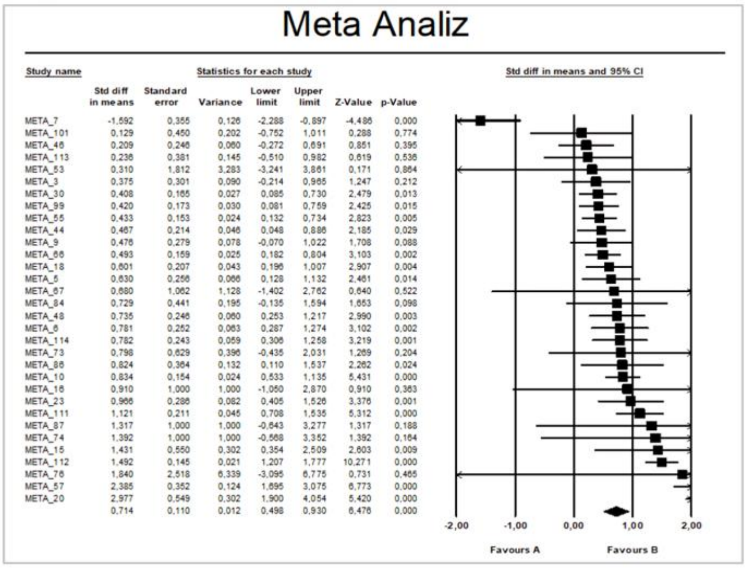
In **Table 1**, Cohen's kappa value, which is the ICR, is statistically significant for primary studies published in both Turkish and English ($^*p < 0.05$). The Cohen's kappa value calculated for studies whose publication language is Turkish is ($\kappa = 0.779$), and it is at a significant level ($0.61 < \kappa < 0.80$) according to Landis and Koch (1977); Fleiss et al. (1981), on the other hand, there is a perfect fit ($0.75 < \kappa$). The Cohen's kappa value calculated for studies whose language is English is ($\kappa = 0.688$) and is moderate ($0.41 < \kappa < 0.60$) according to Landis and Koch (1977). According to Fleiss et al. (1981), there is a good level of agreement ($0.40 < \kappa < 0.75$).

For coder reliability, it was important that those performing the coding had experience with meta-analysis. Two researchers were selected: one, an associate professor in science education, and the other, a professor in computer and instructional technologies. A face-to-face meeting was held with the parallel coders to reach a consensus on the coding. The characteristics of the primary studies included in the study are summarized in **Appendix A**. Citations of selected primary studies are marked with "*" in the bibliography.

Appendix A shows that 69% of the primary studies were published as journal articles, and 84% were published in English. While the research method of 38% of the primary studies was quantitative, 63% of them were mixed, but the data on computational thinking were collected quantitatively. The research design in **Table 2** where 6% of the primary studies are real experimental and 56% are semi-experimental. Of the

Table 2. Heterogeneity analysis for primary studies

Heterogeneity				Tau square			
Q	SD (Q)	p	I^2	τ^2 (tau square)	Standard error	Variance	τ (tau)
141.309	31	0.000	78.062	0.243	0.098	0.010	0.,493



Study name	Statistics for each study						
	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
META_7	-1,592	0,355	0,126	-2,288	-0,897	-4,486	0,000
META_101	0,129	0,450	0,202	-0,752	1,011	0,288	0,774
META_46	0,209	0,246	0,060	-0,272	0,691	0,851	0,395
META_113	0,236	0,381	0,145	-0,510	0,982	0,619	0,536
META_53	0,310	1,812	3,283	-3,241	3,861	0,171	0,864
META_3	0,375	0,301	0,090	-0,214	0,965	1,247	0,212
META_30	0,408	0,165	0,027	0,085	0,730	2,479	0,013
META_99	0,420	0,173	0,030	0,081	0,759	2,425	0,015
META_55	0,433	0,153	0,024	0,132	0,734	2,823	0,005
META_44	0,467	0,214	0,046	0,048	0,886	2,185	0,029
META_9	0,476	0,279	0,078	-0,070	1,022	1,706	0,088
META_66	0,493	0,159	0,025	0,182	0,804	3,103	0,002
META_18	0,601	0,207	0,043	0,196	1,007	2,907	0,004
META_5	0,630	0,256	0,066	0,128	1,132	2,461	0,014
META_67	0,680	1,062	1,128	-1,402	2,762	0,640	0,522
META_84	0,729	0,441	0,195	-0,135	1,594	1,653	0,098
META_48	0,735	0,246	0,060	0,253	1,217	2,990	0,003
META_6	0,781	0,252	0,063	0,287	1,274	3,102	0,002
META_114	0,782	0,243	0,059	0,306	1,258	3,219	0,001
META_73	0,798	0,629	0,396	-0,435	2,031	1,269	0,204
META_86	0,824	0,364	0,132	0,110	1,537	2,262	0,024
META_10	0,834	0,154	0,024	0,533	1,135	5,431	0,000
META_16	0,910	1,000	1,000	-1,060	2,870	0,910	0,363
META_23	0,966	0,286	0,082	0,405	1,526	3,376	0,001
META_111	1,121	0,211	0,045	0,708	1,535	5,312	0,000
META_87	1,317	1,000	1,000	-0,643	3,277	1,317	0,188
META_74	1,392	1,000	1,000	-0,568	3,352	1,392	0,164
META_15	1,431	0,550	0,302	0,354	2,509	2,603	0,009
META_112	1,492	0,145	0,021	1,207	1,777	10,271	0,000
META_76	1,840	2,518	6,339	-3,095	6,775	0,731	0,465
META_57	2,385	0,352	0,124	1,695	3,075	6,773	0,000
META_20	2,977	0,549	0,302	1,900	4,054	5,420	0,000
	0,714	0,110	0,012	0,498	0,930	6,476	0,000

Figure 3. Forest plot for the calculated effect sizes (Kıyıcı, 2022)

primary studies, 38% of which were in the experimental-control group pre-/post-test model, and 34% of the primary studies were planned in the single group pre-/post-test model. At the school level of the research group, 53% are in the 5th and 8th grade range, and 22% are in the 9th and 12th grade range. 41% of primary studies reported pilot studies.

RESULTS

Meta-analysis results will be examined under five sub-headings: heterogeneity analysis, overall moderate effect size, power analysis, publication bias and moderator analyses.

Heterogeneity Analysis

In this study, the effect sizes of the studies on computational thinking in the field of science education between 2006-2021 were combined and a random effects model was chosen in order not to ignore the inter-study variability (Borenstein et al., 2009).

Table 2 shows the results of Cochran Q, Higgins I^2 and T^2 analyzes and summarizes the heterogeneity analysis data for the primary studies of this study. For the Q statistic, the p-value is < 0.01, the I2 value is 78.062 % and the tau-square is 0.243. The calculated Q = 141,309, and this value is greater than the chi-square value (44.495), which corresponds to 31 degrees of freedom. This shows that the effect sizes have a heterogeneous distribution according to the fixed effects model. According to **Table 2**, the I^2 value being 78,062 shows a very significant level of heterogeneity (Borenstein et al., 2009), and it means that 78.062 % of the observed variance comes from actual differences between studies and can therefore potentially be explained by study-level covariates. The I^2 value of this study indicates a high level of heterogeneity. Therefore, it was decided to use the random effects model in the study. If we had an infinitely large study sample, the estimated value in each study would be the true effect and the variance of these effects would be tau-square (τ^2 = 0.243). However, it is difficult to interpret the numerical value of tau-square (τ^2).

In the forest plot in **Figure 3**, the effect size is squared and the horizontal lines extending from the middle of the squares show the confidence interval of the research. The degree of heterogeneity can also be evaluated visually with forest plots. The less the intersection of the point estimates of the studies in the confidence intervals, the higher the heterogeneity. According to **Figure 3**, heterogeneity is seen.

Table 3. Overall effect size details

Model	df	Overall effect size	Variance	Absence hypothesis test (2-tailed)		Standard error	Confidence interval %95	
				Z	p		Lower bound	Upper bound
Fixed effects model	31	0.697	0.002	15.293	0.000	0.046	0.608	0.787
Random effects model	31	0.714	0.012	6.476	0.000	0.110	0.498	0.930

Table 4. Effect size values of primary studies and impact levels according to Cohen et al. (2007)

Research code	Cohen's d	Impact level
META_20 (Walliman, 2015)	2.98	High
META_57 (Yin et al., 2020)	2.39	High
META_76 (Basu et al., 2015)	1.84	High
META_112 (Hava & Ünlü, 2021)	1.49	High
META_15 (Tchoubar, 2018)	1.43	High
META_74 (Garneli & Chorianopoulos, 2018)	1.39	High
META_87 (Sırakaya et al., 2020)	1.32	High
META_111 (Batı & Yetişir, 2021)	1.12	High
META_23 (Christensen, 2020)	0.97	High
META_16 (Akşit, 2018)	0.91	High
META_10 (Adsay et al., 2020)	0.83	High
META_86 (Karaahmetoğlu & Korkmaz, 2019)	0.82	High
META_73 (Merkouris & Chorianopoulos, 2018)	0.80	High
META_114 (Ridlo et al., 2021)	0.78	Middle
META_6 (Bolat, 2020)	0.78	Middle
META_48 (Jaipal-Jamani & Angeli, 2017)	0.74	Middle
META_84 (Batı et al., 2018)	0.73	Middle
META_67 (Irgens et al., 2020)	0.68	Middle
META_5 (Akyol, 2020)	0.63	Middle
META_18 (Basu et al., 2017)	0.60	Middle
META_66 (Zhang et al., 2019)	0.49	Low
META_9 (Cakır & Yaman, 2018)	0.48	Low
META_44 (Fakhriyah et al., 2019)	0.47	Low
META_55 (Hutchins et al., 2020)	0.43	Low
META_99 (Ouyang et al., 2018)	0.42	Low
META_30 (Irgens et al., 2020)	0.41	Low
META_3 (Çimentepe, 2019)	0.38	Low
META_53 (Román-González et al., 2017)	0.31	Low
META_113 (Sun et al., 2020)	0.24	Low
META_46 (Mensan et al., 2020)	0.21	Low
META_101 (Leonard et al., 2018)	0.13	Ineffective
META_7 (Uşengül & Bahçeci, 2020)	-1.59	Ineffective

Overall Moderate Effect Size

In the fixed-effects model, the mean of the universe effect sizes is assumed to be zero, while in the random-effects model, it is assumed that the mean of the effect sizes varies from study to study in the universe. Therefore, the standard deviation of the universe effect sizes of all studies is different from zero (Ellis, 2010). In the random effects model, in Table 3 the Z test, in which the hypothesis accepts that the mean effect sizes are zero in the universe, is significant ($p = 0.000$; $p < 0.05$). As a result, the mean effect sizes are not zero in the universe. Table 3 shows the values of the effect size, Sd, variance and hypothesis testing calculated according to the fixed effect and random effects model. The analysis following the random effects model to predict the effectiveness of science education on computational thinking was moderate ($d = 0.714$, $0.50 < d < 0.80$) (Cohen et al., 2007). The lower limit of the 95% confidence interval is 0.498, the upper limit is 0.930; and the overall effect size of 0.714 (Table 3).

In this study, the self-reported values of the studies whose effect size was reported as Cohen's d, the g values of the studies reported as hedge's g were converted to Cohen's d and the r-value of the studies reported as correlational (r) was converted to Cohen's d (Table 4). In addition, the effect size classification was made by Cohen et al. (2007) and reported in the results section. Table 4 shows that there are 13 primary studies with high, 7 moderate, 10 low and 2 negative impact levels.

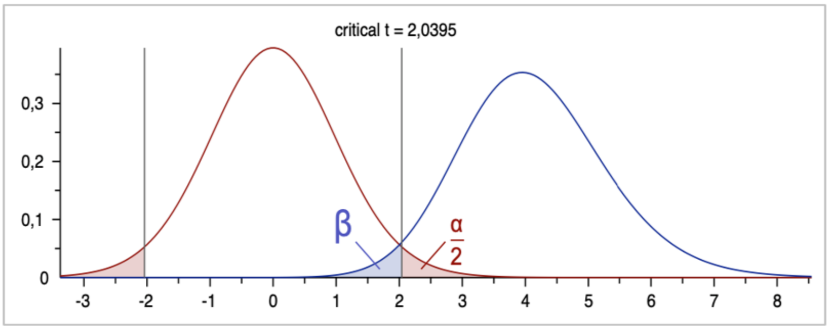


Figure 4. Central and decentralized distributions of the effect size obtained from the research (Kıyıcı, 2022)

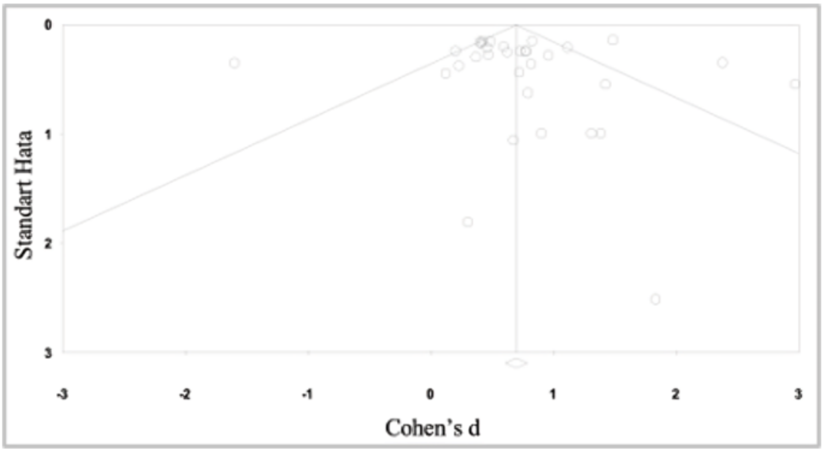


Figure 5. Funnel plot (Kıyıcı, 2022)

In addition, Cohen's d value of one study was negative (**Table 4**). A negative effect size means that the study did not contribute to the overall effect size. When the results of the aforementioned primary research were examined, the post-test score was lower than the pre-test score.

Power Analysis

Power analysis is performed to determine the size of the effect that can be detected with a certain number of studies and whether this detectable effect will be significant (Pigott, 2012).

$$\left.\begin{aligned} \delta &= 0.714 \text{ (effect size)} \\ V_\delta &= 0.012 \text{ (variance)} \\ \alpha &= 0.05 \,\&\, Z = 4.843 \\ c_\alpha &= 1.96 \end{aligned}\right\} \quad \begin{aligned} \lambda^* = \frac{\delta}{\sqrt{V_\delta}} = \frac{0.714}{\sqrt{0.012}} \quad &\longrightarrow \quad \lambda^* = 6.52 \\ p = 1 - \Phi(c_\alpha - \lambda^*) + \Phi(-c_\alpha - \lambda^*) \quad & \\ \mathrm{p} = 1 - \Phi(1.96 - 6.52) + \Phi(-1.96 - 6.52) \quad &\longrightarrow \quad \mathrm{p} = 0.999 \end{aligned}$$

The calculated statistical power shows that the effect size of 32 studies included in the study has a high statistical power.

Figure 4 shows the central and decentralized distributions of the effect size detected in the study. High statistical power was calculated for the effect size value of 32 studies with the G*Power analysis program (p = 0.97). **Figure 4** shows the central and decentralized distributions of the effect size detected in the study.

Publication Bias

To assess the risk of publication bias in the results of this meta-analysis, Funnel scatterplot, Egger regression, Duval and Tweedie (2000) trim and fill, Rosenthal's (1979) error protection number, and Orwin protection number analyzes were performed.

The funnel plot is shown in **Figure 5**. The majority of studies are included in the funnel which means that primary studies are clustered around the average effect size. The tiny circles in the graph represent the primary studies included in the research (Hunter & Schmidt, 2004; Sterne & Egger, 2005). Since the base of

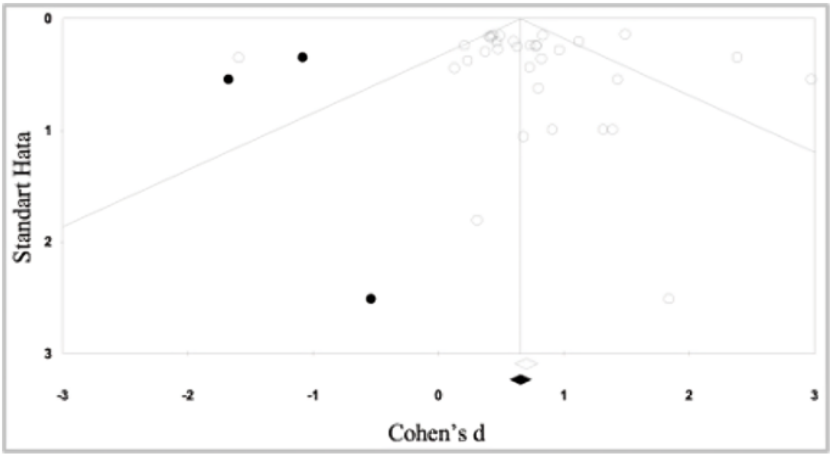


Figure 6. Funnel plot for trim and fill (Kıyıcı, 2022)

Table 5. Results of Duval and Tweedie (2000) trim and fill

	Trimmed studies	Overall effect size	Confidence interval %95	
			Lower bound	Upper bound
Observed values		0.734	0.499	0.969
Estimated values	10	0.447	0.183	0.709

the "funnel" shape of the graph expands as the standard error increases, it is a graph obtained from unbiased data. The funnel plot in **Figure 5** is a scatter plot, one point for each study included in the meta-analysis.

In the graph in **Figure 6**, circles represent primary studies, and black circles represent studies that, if done, will prevent publication bias.

The answer to the question of what research should have been done so that there is no publication bias in the research can be given by the method known as trim and fill (Becker, 2005). The white rectangle at the bottom of the graph in **Figure 5** shows the overall effect of the original study at the 95% confidence interval, and the black one shows the overall effect obtained by adjusting for publication bias (after the trim and fill application).

According to Duval and Tweedie (2000) trim and fill analysis, while the average effect size determined for primary studies was 0.734, if 10 peer studies were added, the estimated effect size value was calculated as 0.447.

The Egger regression test was not found to be statistically significant ($p = 0.762$, $p < 0.05$) (Egger et al., 1997; Sterne & Egger, 2005). Accordingly, there is no bias in primary studies. The Rosenthal (1979) fail-safe number obtained from this meta-analysis study is 1,530, which is statistically significant since $p < 0.05$ ($p = 0.000$). This figure informs us that for $p > 0.05$ to be included in this meta-analysis, 1,530 studies with a zero effect size value should be conducted in the opposite direction of the studies included in this meta-analysis (Borenstein et al., 2019; Rosenthal, 1979). To control the robustness of the results of the meta-analysis against publication bias, the ratio between the calculated error protection number (N) and the number of primary studies in the meta-analysis (k) was calculated as $1{,}530/(5\times32+10) = 9$ (Mullen et al., 2001). According to the results of the Orwin protection number test, to reduce the average effect size of studies on computational thinking in science education below 0.05, 415 studies with an average effect size of 0.000 should be done or 415 studies should be neglected (Orwin & Boruch, 1982). Begg and Mazumdar (1994) rank correlations tests were applied to statistical tests that the study sample was not biased. According to the analysis results in **Table 5**, the primary studies included in this meta-analysis study were not selected biased (tau = 0.14, $p > .05$) (Begg & Mazumdar, 1994).

Moderator Analyses

Moderator variables were determined to investigate how the practices related to science teaching changed the effect of computational thinking. For moderator variable analysis, the values of each variable were categorized into theoretically and statistically comparable codes. Variables that were determined to be moderator as a result of the literature review: publication type, publication language, country of application, pilot study status, year of publication, research method, research design, research model, sampling method of the study group, demographic structure of the study group, school of the study group, type, research area of the application, duration of the application, the person making the application, using a computer, coding, robotics, algorithm, and flipped classroom method in the application, type of measurement tool, and the person who developed the measurement tools.

The moderator analysis results made by the mixed effects model are presented in **Appendix A** and **Appendix B**. **Appendix C** shows the heterogeneity analysis results for subgroups of moderators.

DISCUSSION

In this study, the overall average effect of science education on computational thinking was investigated and the effect of different moderators was determined. This meta-analysis study provided an overview of the evidence that science teaching supports computational thinking. The calculated average effect size value showed that science teaching had a moderate effect on computational thinking, which is consistent with related studies in the literature (Guzdial, 1994; Hambrusch et al., 2009; Jona et al., 2014; NRC, 2000; Pintrich et al., 1991; Redish & Wilson, 1993; Repenning et al., 2010; Sengupta et al., 2013; Sherin, 2001; Wilensky, 1995; Wilensky & Reisman, 2006; Wilensky et al., 2014). In a study on the subject of biological evolution knowledge, which showed an increase in computational thinking knowledge, both evolutionary knowledge change over time and computational thinking over time is important with medium to large effect sizes with univariate analysis of variance (Christensen, 2020). The reporting of different effect sizes on computational thinking in primary studies is an important finding to provide evidence for the impact of science teaching. As a result, effect sizes of all primary studies are not common and moderator variables are effective on effect sizes. The narrow confidence interval of the mean effect size calculated as 0.714 indicates the presence of a small variance for the mean effect size; the wider confidence interval results from the high heterogeneity of the data. The mean effect size values for the random effects model and the fixed effects model are 0.714 and 0.697, respectively. Because of the difference in weighting primary studies, the average effect size estimated with the fixed effects model is usually small. The fixed effects model places less weight on information from studies with small samples and more weight on studies with larger samples to estimate the average effect size. The random effects model, on the other hand, weights primary studies more evenly to estimate the average effect size (Borenstein et al., 2009). Therefore, in this meta-analysis, the average effect size calculated with the random effects model was preferred.

This statistically determined heterogeneity can be due to either sampling error of primary studies or variability between studies. The reason why one of the sources of heterogeneity is thought to be sampling error is the diversity in the samples of primary studies (Huedo-Medina et al., 2006). For example, Basu et al. (2015) and Çimentepe (2019) studied with sixth-grade students, and Batı et al. (2018) worked with eighth-grade students. However, the tests and measurements we use to detect the presence of heterogeneity are not related to the variance value among primary studies, but to the increase in the variance due to heterogeneity (Mittlböck & Heinzl, 2006). Since the Cochran Q statistic calculated in this study is significant, the effect sizes of all primary studies are not common, and the moderator variables are effective on the effect sizes. Another calculated measure of heterogeneity, I2, shows that 78.062% of the observed variance is due to real differences between studies. For this reason, considering that the heterogeneity detected could potentially be explained by covariates at the study level, moderator variable analyzes were performed.

The statistical power calculation made to conduct research free of errors shows that the overall effect size of the 32 studies included in the study has a high statistical power. The high statistical power indicated that the primary studies were sufficient to determine the overall effect level and heterogeneity. Another contribution of the power analysis is that it shows that the probability of rejection is very low by not accepting the H_1 hypothesis, which is true even if the findings support it.

While evaluating the reliability of the results of the meta-analysis, attention was paid to factors that may affect the estimation of the difference resulting from the practices in the primary studies. Studies with biased interventions and for which relevant data were not available even if they met other inclusion criteria were excluded from the meta-analysis so that bias in the selection of primary studies would not affect the research results. Since it is difficult to identify the sources of bias at this stage while determining the inclusion and exclusion criteria, the meta-analysis studies in the literature were examined and expert opinions were taken. However, the bias in primary research is also due to publication bias at the time of publication. The publication tendency of studies with a statistically significant relationship is higher than those with no significant difference. In the 95% confidence interval, the upper limit is 0.930 and the lower limit is 0.498, and the effect size value is 0.714, and it is at a medium ($0,5 \leq d < 0,80$) level (Cohen, 1988). Therefore, the calculated overall effect size value is 95% accurate.

In the Egger Regression analysis, there was no publication bias among the primary studies. On the other hand, the fact that Rosenthal's (1979) error protection number is significant indicates that there is no publication bias in primary studies. For the significance to disappear as a result of the meta-analysis, 1530 studies with a zero effect size value should be done or neglected. The Rosenthal (1979) error protection number is 47 times the number of included studies. The 32 studies included in the meta-analysis are all studies for this research question that can be accessed from all studies according to inclusion criteria. Apart from these, there is no possibility of reaching 1,530 studies, and this result is an indication that there is no publication bias in the meta-analysis study. The Orwin protection number is 415, and if 415 studies can be accessed apart from the 32 primary studies included in this study, the effect size obtained will be reduced to the level of indifference ($p < 0.05$). The Orwin's safe N number is more than 12 times the number of studies included in the meta-analysis can be considered as an indication of the absence of publication bias. The applied Begg and Mazumdar (1994) rank correlations were found to be insignificant. As a result, since the sample selection is not biased, it does not cause any publication bias. Analysis of publication bias showed that publication bias did not affect the effect size in computational thinking studies in science education. This is an important finding showing that the evidence for the impact of computational thinking in science education is strong, although previous studies have reported varying levels of effect size (Christensen, 2020; Leonard et al., 2018; Ridlo et al., 2021).

One of the variables regulating the relationship between science education and computational thinking is not the type of publication, it also overlaps with the findings of different studies in the literature.

The publication type (article and thesis) did not change the effect of augmented reality on academic achievement. Similarly, in the study investigating the effect of portfolio usage on academic success, the publication type (master's thesis, doctoral thesis, article, and conference document) did not have a moderator effect (Akçayır, 2018; Başol & Erbay, 2017; Batdı, 2014; Shin & Kim, 2013; Şen & Yılmaz, 2013). The finding that publication type was not a significant moderator in this meta-analysis warrants more careful consideration than simply stating that it was not significant. There can be several reasons for this result. First, the included studies sufficiently varied to show clear differences; most were experimental or quasi-experimental studies published in peer-reviewed journals, which limits variation in publication type as a category. If a moderator variable varies little across studies, it is harder to detect moderation effects, which may lead to non-significant results even when a real effect exists (Borenstein et al., 2009). Second, in theory, the effectiveness of computational thinking interventions in science education may depend mostly on teaching methods and context—such as how instruction is delivered, how long the intervention lasts, or how well it aligns with curriculum goals—rather than on how the findings are published. This suggests that the link between computational thinking and science achievement is strong across publication types, consistent with the steady effect sizes observed in the studies.

The relevant data indicate that the average effect size of the pilot study is statistically insignificant. This meaninglessness shows that the effectiveness of science education on computational thinking skills did not affect the conduct of the pilot study. However, pilot applications in educational research increase the strengths of the studies by providing a critical perspective and supporting the validity of their results (Gudmundsdottir & Brock-Utne, 2010). The pilot study provides to control the research focus, questions, and method by gaining foresight about the actual research. Therefore, reporting pilot studies that contribute to the scientific validity and power of primary studies is also important for meta-analysis studies. However, the limited use of pilot studies in the primary research of this study may have revealed this inconsistency.

For the interaction between science education and the level of computational thinking, the average effect size of the publication year moderator was not statistically significant. This is an indication that there may be uncontrollable problems in these studies as well as the low number of studies in subgroups. When meta-analyses across different topics in the literature are examined, the year of publication is not a moderator variable (Dağyar & Demirel, 2015; Tutal, 2019). and statistically significant for some subgroups and insignificant for others (Othman, 1996).

Considering the time required for the recognition of its relationship with science education after 2006, which is widely accepted as the starting point of computational thinking research worldwide, it becomes more understandable that the average effect sizes of the subgroups are significant as 2021 approaches.

When Barr and Stephenson's (2011) framework for integrating computational thinking into science education is examined, the fields of physics and chemistry are well-suited for developing computational thinking sub-skills in terms of knowledge qualities and learning processes. However, while science (n = 9), STEM (n = 15), and physics-chemistry-biology (n = 6) fields are thought to be suitable for developing computational thinking sub-skills, the statistical results indicate the opposite.

In the moderator variable analysis, the application period did not regulate the effect of science education on computational thinking skills. Although the average effect sizes across application duration subgroups differ, there is no statistically significant difference among them (Armağan, 2011; Ayaz, 2014; Üstün, 2012). The number of primary studies that determined the effect of science education on computational thinking was limited. Therefore, it is clear that more primary studies are needed to conduct more comprehensive meta-analyses. It is important to determine policies, assess the effectiveness of teaching methods and techniques, and implement these decisions, as indicated by meta-analyses in education in general and science education in particular.

Limitations

This study is limited by the constraints inherent in the meta-analysis method. Researchers usually report expected results, which can lead to the disregard of unexpected results, even if they are statistically significant. As a result, this reduces the validity of the meta-analysis. The researcher decides on the variables and criteria for selecting studies to be included in the research. Therefore, the objectivity of the meta-analysis method is low (Armağan, 2011). Clearly stating criteria and decisions is critical to ensuring the reliability and validity of any study.

First, the scope of the literature search was confined to studies indexed in specific databases—including the Council of Higher Education Thesis Catalog, ULAKBİM, ProQuest Digital Dissertations, ProQuest EBSCOhost, ERIC, ScienceDirect, Web of Science, Taylor & Francis Online, Scopus, and JSTOR—and to studies published between January 2006 and April 2021. Studies published outside this timeframe or indexed in other repositories were not included. While multiple databases were searched, the inclusion of gray literature remained limited, which may introduce a degree of publication bias risk despite the favorable results of the bias analyses conducted in this study.

Other limitation, the literature search was restricted to studies containing the keywords "STEM," "Science Education," "Science Teaching," "Science Learning," and "Computational Thinking." This keyword selection, though systematic, may have excluded relevant studies that used alternative terminology or conceptual framings of computational thinking in science contexts. Future searches might consider broader or supplementary keyword sets to capture a more comprehensive body of literature.

Like all meta-analyses, this study reflects decisions made by the researcher regarding inclusion and exclusion criteria, coding categories, and the operationalization of moderator variables. These decisions introduce an inherent degree of subjectivity. To mitigate this, all criteria were defined a priori and applied consistently throughout the screening and coding process. Nevertheless, readers should interpret the findings in light of these methodological choices, particularly the exclusion of studies that did not report sufficient statistical data to calculate effect sizes.

Finally, the moderator analyses were limited to the 23 variables captured in the coding form, which, in turn, were constrained by the information reported in the primary studies. Several theoretically relevant variables—such as teacher training level, implementation fidelity, or the specific computational thinking components targeted—were frequently underreported in the primary studies and therefore could not be examined as moderators. This represents not only a limitation of the current study but also a gap in the primary literature that future researchers should address by adopting more standardized reporting practices. Of course, the data in the coding form is limited to the information reported in the studies included in the meta-analysis. It also covers research from January 2006 to April 2021. Literature published before or after these dates was not included in the review. Although different search tools have been used, the amount of gray literature included in the research is limited. Hence, this may pose a threat to publication bias.

Implications

In this study, a medium effect size was determined on the computational thinking level of science education activities. However, there are gaps in the literature regarding the effects of activities developed for computational thinking on science achievement or attitudes. In addition, we have noticed that research on computational thinking is mostly conducted in computer-related fields. However, it attaches particular importance to its integration with science, as computational thinking is considered the connective tissue between computer science and science.

Computational thinking is an umbrella concept that contains various skills (Ibılı et al., 2020). Studies on computational thinking mostly focus on abstraction, problem-solving, and algorithmic thinking components (Kalelioglu et al., 2016). However, there is a lack of studies in the literature that examine the relationships among these components in a multifaceted manner.

In the present study, we found that 8 out of 32 primary studies used random sampling. Although it is often stated that the sample was randomly selected, in practice, as Delen and Sen (2022) noted, randomization is usually applied at the group level rather than the student level. In other words, student clusters are randomly assigned to either the experimental or the control group. Whereas randomization should ideally be used to evenly distribute students' characteristics. When choosing a sampling method in future research for data with a higher effect size and stronger against publication bias (Silver & Kelsay, 2021), care should be taken.

Author contributions: GK: conceptualization, investigation, methodology, software, writing – original draft; **HY:** data curation, formal analysis, supervision, validation, writing – review & editing. Both authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: This study is a meta-analysis and does not involve the collection of primary data from human participants. Therefore, ethical approval and participant consent were not required.

AI statement: No generative AI or AI-based tools were used during this study.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

*Adsay, C., Korkmaz, Ö., Çakır, R., & Erdoğmuş, F. U. (2020). Ortaokul öğrencilerinin blok temelli kodlama eğitimine dönük öz-yeterlik agı düzeyleri, STEM ve bilgi işlemsel düşünme beceri düzeyleri [Middle school students' self-efficacy network levels towards block-based coding education, STEM and computational thinking skills]. *Eğitim Teknolojisi Kuram ve Uygulama, 10*(2), 469-489. https://doi.org/10.17943/etku.696224

Akçayır, G. (2018). *Artırılmış gerçekliğin eğitimde etkisinin incelenmesi: Meta-analiz ve sistematik kaynak taraması araştırması* [Examining the impact of augmented reality in education: A meta-analysis and systematic literature review study] [PhD thesis, Gazi University].

*Akşit, O. (2018). *Enhancing science learning through computational thinking and modeling in middle school classrooms: A mixed methods study* [PhD thesis, North Carolina State University].

*Akyol, B. E. (2020). *STEM etkinliklerinin fen bilgisi öğretmen adaylarının bilgi işlemsel, eeştirel, yaratıcı düşünme ve problem çözme becerilerine etkisi* [The impact of STEM activities on the computational, critical, creative thinking, and problem-solving skills of prospective science teachers] [PhD thesis, Erciyes University].

Armağan, F. Ö. (2011). *Kavramsal değişim metinlerinin etkililiği: Meta analiz çalışması* [The effectiveness of conceptual change texts: A meta-analysis study] [Gazi University].

Ayaz, M. F. (2014). *Proje tabanlı öğrenme yaklaşımının öğrencilerin fen derslerindeki akademik başarılarına ve fen derslerine yönelik tutumlarına etkisi: Bir meta-analiz çalışması* [The effect of project-based learning approach on students' academic achievement and attitudes towards science courses: A meta-analysis study] [PhD thesis, Dicle University]. https://doi.org/10.12780/uusbd.63667

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *ACM Inroads, 2*(1), 48-54. https://doi.org/10.1145/1929887.1929905

*Basu, S., Biswas, G., & Kinnebrew, J. S. (2017). Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. *User Modeling and User-Adapted Interaction, 27*(1), 5-53. https://doi.org/10.1007/s11257-017-9187-0

Basu, S., Biswas, G., Kinnebrew, J., & Rafi, T. (2015). Relations between modeling behavior and learning in a computational thinking based science learning environment. In *Proceedings of the 23rd International Conference on Computers in Education* (pp. 184-189).

Başol, G., & Erbay, Ş. (2017). Portfolyo kullanımının akademik başarıya etkisi: Bir meta analiz çalışması [The impact of portfolio use on academic achievement: A meta-analysis study]. *Hacettepe Eğitim Dergisi, 32*(2), 396-412. https://doi.org/10.16986/HUJE.2017025320

Batdı, V. (2014). Kavam haritası tekniği ile geleneksel öğrenme yönteminin kullanılmasının öğrencilerin başaroları, bilgilerinin kalıcılığı ve tutumlarına etkisi: Bir meta-analiz çalışması [The effect of using concept mapping techniques and traditional learning methods on students' achievement, retention of knowledge, and attitudes: A meta-analysis study]. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (42), 93-102.

*Batı, K., & Yetişir, M. İ. (2021). Examination of Turkish middle school STEM teachers' knowledge about computational thinking and views regarding information and communications technology. *Computers in the Schools, 38*(1), 57-73. https://doi.org/10.1080/07380569.2021.1882206

Batı, K., Yetişir, M. İ., Çalışkan, I., Güneş, G., & Saçan, E. G. (2018). Teaching the concept of time: A steam-based program on computational thinking in science education. *Cogent Education, 5*(1), Article 1507306. https://doi.org/10.1080/2331186X.2018.1507306

Becker, B. J. (2005). Failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 111-126). John Wiley & Sons.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088-1101. https://doi.org/10.2307/2533446

*Bolat, Y. İ. (2020). *STEM temelli matematik etkinliklerinin problem çözme ve bilgi işlemsel düşünme becerisi ile STEM alanlarına olan ilgiye katkılarının araştırılması* [Investigation of the contributions of STEM-based mathematics activities to problem solving and computational thinking skills and interest in STEM fields] [PhD thesis, Atatürk University].

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

*Cakır, E., & Yaman, S. (2018). Ters yüz sınıf modelinin öğrencilerin fen başarısı ve bilgisayarca düşünme becerileri üzerine etkisi [The effect of the flipped classroom model on students' science achievement and computational thinking skills]. *Gazi University Journal of Gazi Education Faculty, 38*(1), 75-99.

Card, N. A. (2012). Applied meta-analysis for social science research. In D. A. Kenny, & T. D. Little (Eds.), *Applied meta-analysis for social science research*. The Guilford Press.

*Christensen, D. M. (2020). *Learning biological evolution through computational thinking* [PhD thesis, Temple University].

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. Routledge. https://doi.org/10.4324/9780203029053

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological testing and assessment*. In Guidebook for clinical pychology interns. McGraw Hill. https://doi.org/10.1007/978-1-4899-0222-1_8

Cömert, S., & Kıyıcı, F. B. (2013). Fen bilgisi öğretiminde oluşturmacu yaklaşım uygulamaısnın akademik başarıya etkisinin belirlenmesi [Determining the effect of applying the constructivist approach in science education on academic achievement]. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi,* (7), 151-162.

*Çimentepe, H. E. (2019). *STEM etkinliklerinin akademik başarı, bilimsel süreç becerileri ve bilgisayarca düşünme becerilerine etkisi* [The impact of STEM activities on academic achievement, scientific process skills, and computational thinking skills] [Master's thesis, Niğde Ömer Halisdemir University].

Dağyar, M., & Demirel, M. (2015). Effects of problem-based learning on academic achievement: A meta-analysis study. *Eğitim ve Bilim, 40*(181), 139-174. https://doi.org/10.15390/EB.2015.4429

Delen, I., & Sen, S. (2022). Effect of design-based learning on achievement in K-12 education: A meta-analysis. *Journal of Research in Science Teaching, 60*(2), 330-356. https://doi.org/10.1002/tea.21800

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89-98. https://doi.org/10.1080/01621459.2000.10473905

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

Ellis, P. D. (2010). *The essential guide to effect sizes*. Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

*Fakhriyah, F., Masfuah, S., & Mardapi, D. (2019). Developing scientific literacy-based teaching materials to improve students' computational thinking skills. *Jurnal Pendidikan IPA Indonesia, 8*(4), 482-491. https://doi.org/10.15294/jpii.v8i4.19259

Fleiss, J., Levin, B., & Paik, M. (1981). *Statistical methods for rates and proportions*. Wiley.

*Garneli, V., & Chorianopoulos, K. (2018). Programming video games and simulations in science education: exploring computational thinking through code analysis. *Interactive Learning Environments, 26*(3), 386-401. https://doi.org/10.1080/10494820.2017.1337036

Göktaş, E. (2017). *Eğitim politikası bağlamında işbirlikli öğrenme ve geleneksel öğrenme yöntemlerinin başarı ve tutuma etkisinin meta analitik biçimde incelenmesi* [A meta-analytic examination of the impact of collaborative learning and traditional learning methods on achievement and attitude within the context of education policy] [PhD thesis, Marmara University].

Grover, S., & Pea, R. (2017). Computational thinking: A competency whose time has come. In *Computer science education: Perspectives on teaching and learning* (pp. 19-38). Bloomsbury Publishing. https://doi.org/10.5040/9781350057142.ch-003

Gudmundsdottir, G. B., & Brock-Utne, B. (2010). An exploration of the importance of piloting and access as action research. *Educational Action Research, 18*(3), 359-372. https://doi.org/10.1080/09650792.2010.499815

Guzdial, M. (1994). Software-realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments, 4*(1), 1-44. https://doi.org/10.1080/1049482940040101

Hambrusch, S., Hoffmann, C., Korb, J. T., Haugan, M., & Hosking, A. L. (2009). A multidisciplinary approach towards computational thinking for science majors. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education* (pp. 183-187). ACM. https://doi.org/10.1145/1508865.1508931

*Hava, K., & Ünlü, Z. (2021). Investigation of the relationship between middle school students' computational thinking skills and their STEM career interest and attitudes toward inquiry. *Journal of Science Education and Technology, 30*, 484-495. https://doi.org/10.1007/s10956-020-09892-y

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hsu, T. C., Chang, S. C., & Hung, Y. T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers and Education, 126*, 296-310. https://doi.org/10.1016/j.compedu.2018.07.004

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11*(2), 193-206. https://doi.org/10.1037/1082-989X.11.2.193

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta analysis*. SAGE. https://doi.org/10.4135/9781412985031

*Hutchins, N. M., Biswas, G., Maróti, M., Lédeczi, Á., Grover, S., Wolf, R., Blair, K. P., Chin, D., Conlin, L., Basu, S., & McElhaney, K. (2020). C2STEM: A system for synergistic learning of physics and computational thinking. *Journal of Science Education and Technology, 29*(1), 83-100. https://doi.org/10.1007/s10956-019-09804-9

Ibılı, E., Günbatar, M. S., & Sırakaya, M. (2020). Bilgi-işlemsel düşünme becerilerinin incelenmesi: Meslek liseleri örneklemi [Examination of computational thinking skills: A sample of vocational high schools]. *Kastamonu Education Journal, 28*(2), 1067-1078. https://doi.org/10.24106/kefdergi.683577

*Irgens, G. A., Dabholkar, S., Bain, C., Woods, P., Hall, K., Swanson, H., Horn, M., & Wilensky, U. (2020). Modeling and measuring high school students' computational thinking practices in science. *Journal of Science Education and Technology, 29*(1), 137-161. https://doi.org/10.1007/s10956-020-09811-1

*Jaipal-Jamani, K., & Angeli, C. (2017). Effect of robotics on elementary preservice teachers' self-efficacy, science learning, and computational thinking. *Journal of Science Education and Technology, 26*(2), 175-192. https://doi.org/10.1007/s10956-016-9663-z

Jona, K., Wilensky, U., Trouille, L., Horn, M., Orton, K., Weintrop, D., & Beheshti, E. (2014). Embedding computational thinking in science , technology , engineering , and math (CT-STEM). *Future Directions in Computer Science Education Summit Meeting*. http://ccl.sesp.northwestern.edu/papers/2014/OrtonKaiNorthwestern-1.pdf

Kalelioglu, F., Gulbahar, Y., & Kukul, V. (2016). A framework for computational thinking based on a systematic research review. *Baltic Journal Modern Computing, 4*(3), 583-596.

*Karaahmetoğlu, K., & Korkmaz, Ö. (2019). The effect of project-based Arduino educational robot applications on students' computational thinking skills and their perception of basic STEM skill levels. *Participatory Educational Research, 6*(2), 1-14. https://doi.org/10.17275/per.19.8.6.2

Kıyıcı, G. (2022). *Fen eğitiminde hesaplamalı düşünme araştırmalarının incelenmesi: Bir meta-analiz çalışması* [Examining research on computational thinking in science education: A meta-analysis study] [PhD thesis, Gazi University].

Korkmaz, Ö., Çakır, R., & Özden, M. Y. (2015). Bilgisayarca düşünme beceri düzeyleri ölçeğinin (BDBD) ortaokul düzeyine uyarlanması [Adaptation of the computational thinking skills level scale (CTSL) to the middle school level]. *Gazi Eğitim Bilimleri Dergisi, 1*(2), 143-162.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), Article 159. https://doi.org/10.2307/2529310

Lei, H., Chiu, M. M., Li, F., Wang, X., & Geng, Y. (2020). Computational thinking and academic achievement: A meta-analysis among students. *Children and Youth Services Review, 118*, Article 105439. https://doi.org/10.1016/j.childyouth.2020.105439

*Leonard, J., Mitchell, M., Barnes-Johnson, J., Unertl, A., Outka-Hill, J., Robinson, R., & Hester-Croff, C. (2018). Preparing teachers to engage rural students in computational thinking through robotics, game design, and culturally responsive teaching. *Journal of Teacher Education, 69*(4), 386-407. https://doi.org/10.1177/0022487117732317

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine, 6*(7), Article e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lipsey, M. W., & Wilson, D. (2001). Practical meta analysis overview. In L. Bickman, & D. Rog (Eds.), *Applied social research methods series*. SAGE.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. Oxford University Press.

Luo, F., Antonenko, P. D., & Davis, E. C. (2020). Exploring the evolution of two girls ' conceptions and practices in computational thinking in science. *Computers & Education, 146*, Article 103759. https://doi.org/10.1016/j.compedu.2019.103759

*Mensan, T., Osman, K., & Majid, N. A. A. (2020). Development and validation of unplugged activity of computational thinking in science module to integrate computational thinking in primary science education. *Science Education International, 31*(2), 142-149. https://doi.org/10.33828/sei.v31.i2.2

*Merkouris, A., & Chorianopoulos, K. (2018). Programming touch and full-body interaction with a remotely controlled robot in a secondary education STEM course. In *Proceedings of 22nd Pan-Hellenic Conference on Informatics*. https://doi.org/10.1145/3291533.3291537

Mittlböck, M., & Heinzl, H. (2006). A simulation study comparing properties of heterogeneity measures in meta-analyses. *Statistics in Medicine, 25*(24), 4321-4333. https://doi.org/10.1002/sim.2692

Mullen, B., Muellerleile, P., & Bryant, B. (2001). Cumulative meta-analysis: A consideration of indicators of sufficiency and stability. *Personality and Social Psychology Bulletin, 27*(11), 1450-1462. https://doi.org/10.1177/01461672012711006

NRC. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. The National Academies Press. https://doi.org/10.17226/9596

Orwin, R. G., & Boruch, R. F. (1982). RRT meets RDD: Statistical strategies for assuring response privacy in telephone surveys. *The Public Opinion Quarterly, 46*(4), 560-571. https://doi.org/10.1086/268752

Othman, N. (1996). *The effects of cooperative learning and traditional mathematics instruction in grades k-12: A meta-analysis of findings* [PhD thesis, West Virginia University].

*Ouyang, Y., Hayden, K. L., & Remold, J. (2018). Introducing computational thinking through non-programming science activities. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 308-313). ACM. https://doi.org/10.1145/3159450.3159520

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Perkins, D. N ., & Simmons, R. (1988). Patterns of misunderstanding: An integrative model for science, math, and programming. *Review of Educational Research, 58*(3), 303-326. https://doi.org/10.3102/00346543058003303

Pigott, T. D. (2012). *Advances in meta-analysis*. Springer. https://doi.org/10.1007/978-1-4614-2278-5

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire [MSLQ]*. APA PsycTests. https://doi.org/10.1037/t09161-000

Redish, E. F., & Wilson, J. M. (1993). Student programming in the introductory physics course: M.U.P.P.E.T. *American Journal of Physics, 61*(3), 222-232. https://doi.org/10.1119/1.17295

Repenning, A., Webb, D., & Ioannidou, A. (2010). Scalable game design and the development of a checklist for getting computational thinking into public schools. In *Proceedings of the 41st ACM Technical Symposium on Computer Science Education* (pp. 265-269). ACM. https://doi.org/10.1145/1734263.1734357

*Ridlo, Z. R., Indrawati, Afafa, L., Bahri, S., Kamila, I. S., & Rusdianto. (2021). The effectiveness of research-based learning model of teaching integrated with computer simulation in astronomy course in improving student computational thinking skills. *Journal of Physics: Conference Series, 1839*, Article 012027. https://doi.org/10.1088/1742-6596/1839/1/012027

Rohaeti, E., & Huda, N. (2025). Assessing computational thinking skills of science and mathematics upper-secondary school students. *European Journal of Science and Mathematics Education, 13*(4), 289-303. https://doi.org/10.30935/scimath/17248

*Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior, 72*, 678-691. https://doi.org/10.1016/j.chb.2016.08.047

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. https://doi.org/10.1037/0033-2909.86.3.638

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies, 18*(2), 351-380. https://doi.org/10.1007/s10639-012-9240-x

Sherin, B. L. (2001). A comparison of programming languages and algebraic notation as expressive languages for physics. *International Journal of Computers for Mathematical Learning, 6*(1), 1-61. https://doi.org/10.1023/A:1011434026437

Shin, I. S., & Kim, J. H. (2013). The effect of problem-based learning in nursing education: A meta-analysis. *Advances in Health Sciences Education, 18*(5), 1103-1120. https://doi.org/10.1007/s10459-012-9436-2

*Sırakaya, M., Sırakaya, D. A., & Korkmaz, Ö. (2020). The impact of STEM attitude and thinking style on computational thinking determined via structural equation modeling. *Journal of Science Education and Technology, 29*(4), 561-572. https://doi.org/10.1007/s10956-020-09836-6

Silver, I. A., & Kelsay, J. D. (2021). The moderating effects of population characteristics: A potential biasing factor when employing non-random samples to conduct experimental research. *Journal of Experimental Criminology, 19*, 107-118. https://doi.org/10.1007/s11292-021-09478-7

Sterne, J., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In R. H. R., A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis prevention, assessment and adjustments* (pp. 99-110). John Wiley & Sons Ltd. https://doi.org/10.1002/0470870168.ch6

*Sun, L., Hu, L., Yang, W., Zhou, D., & Wang, X. (2020). STEM learning attitude predicts computational thinking skills among primary school students. *Journal of Computer Assisted Learning, 37*, 346-358. https://doi.org/10.1111/jcal.12493

Şen, Ş., & Yılmaz, A. (2013). İşbirlikçi öğrenmenin kavramsal değişim üzerindeki etkisi: Bir meta analiz çalışması [The impact of collaborative learning on conceptual change: A meta-analysis study]. *Karaelmas Eğitim Bilimleri Dergisi, 1*, 21-32.

*Tchoubar, T. (2018). *eLearning model creating readiness to use global information* [PhD thesis, Stony Brook University].

Tutal, Ö. (2019). *Aktif öğrenme yaklaşımının öğrencilerin akademik başarılarına, derse yönelik tutumlarına ve öğrenilenlerin kalıcılığına etkisi: Bir meta-analiz çalışması* [The effect of active learning approach on students' academic achievement, attitudes towards the course, and retention of learning: A meta-analysis study] [PhD thesis, Dicle University].

*Uşengül, L., & Bahçeci, F. (2020). The effect of Lego Wedo 2.0 Education on academic achievement and attitudes and computational thinking skills of learners toward science. *World Journal of Education, 10*(4). https://doi.org/10.5430/wje.v10n4p83

Üstün, U. (2012). *To what extent is problem-based learning effective as compared to traditional teaching in science education? A meta-analysis study* [PhD thesis, Middle East Technical University].

Vevea, J. L., & Zelinsky, N. A. M. (2019). Evaluating coding decision. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 173-206). SAGE.

Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies, 20*(4), 715-728. https://doi.org/10.1007/s10639-015-9412-6

*Walliman, G. (2015). *Genost: A system for introductory computer science education with a focus on computational thinking* [PhD thesis, Arizona State University].

Wilensky, U. (1995). Paradox, programming, and learning probability: A case study in a connected mathematics framework. *Journal of Mathematical Behavior, 14*(2), 253-280. https://doi.org/10.1016/0732-3123(95)90010-1

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories – An embodied modeling approach. *Cognition and Instruction, 24*(2), 171-209. https://doi.org/10.1207/s1532690xci2402_1

Wilensky, U., Brady, C. E., & Horn, M. S. (2014). Fostering computational literacy in science classrooms. *Communications of the ACM, 57*(8), 24-28. https://doi.org/10.1145/2633031

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35. https://doi.org/10.1145/1118178.1118215

Yavru, Ö., & Gürdal, A. (1998). İlköğretim okullarının 4. ve 5. sınıflarında laboratuvar deneylerinin öğrencilerin mekanik konusundaki başarısına ve kavramları kazanmasına etkisi [The effect of laboratory experiments on students' success and acquisition of concepts in mechanics in 4th and 5th grades of primary schools]. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi, 10*(10), 327-338.

*Yin, Y., Hadad, R., Tang, X., & Lin, Q. (2020). Improving and assessing computational thinking in maker activities: The integration with physics and engineering learning. *Journal of Science Education and Technology, 29*(2), 189-214. https://doi.org/10.1007/s10956-019-09794-8

*Zhang, Y., Wang, J., Bolduc, F., & Murray, W. G. (2019). LP based integration of computing and science education in middle schools. In *Proceedings of the ACM Conference on Global Computing Education* (pp. 44-50). ACM. https://doi.org/10.1145/3300115.3309512

APPENDIX A

Table A1. Primary studies

Fine name	Author name	Publication type	Publication language	Pilot study	Research method	Pattern of research	Research model	Research group
META_10	(Adsay et al., 2020)	Journal article	Turkish	Unavailable	Mix method	Non-experimental	Relational screening model	Middle school 6th, 7th, 8th grade
META_101	(Leonard et al., 2018)	Journal article	English	Available	Mix method	Weak experimental	Single group pre-/post-test	Teachers
META_111	(Batı & Yetişir, 2021)	Journal article	English	Unavailable	Quantitative method	Non-experimental	Relational screening model	Teachers
META_112	(Hava & Ünlü, 2021)	Journal article	English	Unavailable	Quantitative method	Non-experimental	Relational screening model	Middle school
META_113	(Sun et al., 2020)	Journal article	English	Unavailable	Quantitative method	Non-experimental	Relational screening model	Middle school
META_114	(Ridlo et al., 2021)	Journal article	English	Unavailable	Mix method	Semi-experimental	Pre-/post-test model with experimental-control	Undergraduate
META_15	(Tchoubar, 2018)	PhD thesis	English	Available	Mix method	Weak experimental	Single group post-test model	Middle school 6th grade
META_16	(Akşit, 2018)	PhD thesis	English	Unavailable	Mix method	Weak experimental	Single group post-test model	Middle school 7th grade
META_18	(Basu et al., 2017)	Journal article	English	Available	Mix method	Semi-experimental	Pre-/post-test model with experimental-control	Middle school 6 class
META_19	(Walliman, 2015)	Master thesis	English	Unavailable	Mix method	Semi-experimental	Single group pre-/post-test	Undergraduate
META_23	(Christensen, 2020)	PhD thesis	English	Available	Mix method	Semi-experimental	Pre-/post-test model with experimental-control	High school 9th, 10th, 11th, 12th grade
META_3	(Çimentepe, 2019)	Master thesis	Turkish	Available	Quantitative method	Semi-experimental	Pre-/post-test model with experimental-control	Middle school 6th grade
META_30	(Irgens et al., 2020)	Journal article	English	Unavailable	Mix method	Weak experimental	Single group pre-/post-test	High school 12th grade
META_44	(Fakhriyah et al., 2019)	Journal article	English	Available	Quantitative method	Semi-experimental	Post-test model with experimental-control	Undergraduate
META_46	(Mensan et al., 2020)	Journal article	English	Available	Quantitative method	Semi-experimental	Pre-/post-test model with experimental-control	Middle school 5th grade
META_48	(Jaipal-Jamani & Angeli, 2017)	Journal article	English	Available	Quantitative method		Single group pretest/post	Undergraduate
META_5	(Akyol, 2020)	PhD thesis	Turkish	Available	Mix method	Semi-experimental	Pre-/post-test model with experimental-control	Undergraduate
META_53	(Román-González et al., 2017)	Journal article	English	Unavailable	Quantitative method	Semi-experimental	Single group pre-/post-test	Middle school 5th, 6th, 7th, 8th grade & high school 9th & 10th grade
META_55	(Hutchins et al., 2020)	Journal article	English	Available	Mix method	Semi-experimental	Pre-/post-test model with experimental-control	High school
META_57	(Yin et al., 2020)	Journal article	English	Unavailable	Mix method	Weak experimental	Single group pre-/post-test	High school
META_6	(Bolat, 2020)	PhD thesis	Turkish	Available	Mix method	Experimental	Pre-/post-test model with experimental-control	High school 10th grade
META_66	(Zhang et al., 2019)	Journal article	English	Available	Quantitative method	Semi-experimental	Pre-/post-test model with experimental-control	High school 6th & 7th grade
META_67	(Irgens et al., 2020)	Journal article	English	Unavailable	Quantitative method	Experimental	Single group pre-/post-test	High school
META_7	(Uşengül & Bahçeci, 2020)	Journal article	English	Unavailable	Quantitative method	Semi-experimental	Single group pre-/post-test	Middle school 5th grade
META_73	(Merkouris & Chorianopoulos, 2018)	Congress notice	English	Unavailable	Mix method	Semi-experimental	Single group post-test model	Middle school 6th grade
META_74	(Garneli & Chorianopoulos, 2018)	Journal article	English	Unavailable	Mix method	Semi-experimental	Single group pre-/post-test	Middle school 6th grade
META_76	(Basu et al., 2015)	Congress notice	English	Unavailable	Mix method	Weak experimental	Single group pre-/post-test	Middle school 5th grade
META_84	(Batı et al., 2018)	Journal article	English	Unavailable	Mix method	Semi-experimental	Pre-/post-test model with experimental-control	Middle school 8th grade
META_86	(Karaahmetoğlu & Korkmaz, 2019)	Journal article	English	Unavailable	Mix method	Semi-experimental	Pre-/post-test model with experimental-control	Middle school 6th grade
META_87	(Sırakaya et al., 2020)	Journal article	English	Unavailable	Mix method	Non-experimental	Relational screening model	Middle school 5th-8th grades
META_9	(Cakır & Yaman, 2018)	Journal article	Turkish	Available	Quantitative method	Semi-experimental	Pre-/post-test model with experimental-control	Middle school 7th grade
META_99	(Ouyang et al., 2018)	Congress notice	English	Available	Mix method	Weak experimental	Single group pre-/post-test	Mix

APPENDIX B

Table B1. Moderator analyses of outcomes, publication bias, generation, and instructional setting

Variable		Effect size and 95% CI								Heterogeneity		
		N	d	SE	Variance	95% CI		Z	p	Q_B	df	p
Publication type	Article	21	0.873	0.145	0.021	0.589	1.156	6.036	0.000	*3.061	1	0.080
	Others	11	0.466	0.182	0.033	0.110	0.823	2.565	0.010			
Language	Turkish	5	0.002	0.285	0.081	-0.560	0.556	0.07	0.985	7.320	1	0.007
	English	27	0.829	0.115	0.013	0.603	1.055	7.190	0.000			
Country	Turkey	11	0.508	0.204	0.042	0.109	0.908	2.398	0.016	*1.774	2.000	0.412
	USA	13	0.751	0.177	0.031	0.404	1.098	3.920	0.000			
	Other countries	8	0.896	0.216	0.047	0.473	1.319	0.288	0.773			
Pilot study	Not specified	4	0.907	0.334	0.111	0.253	1.561	2.718	0.007	*0.628	2.000	0.731
	Yes	13	0.625	0.175	0.031	0.281	0.969	3.563	0.000			
	No	15	0.745	0.160	0.026	0.431	1.059	4.646	0.000			
Publication year	2021	2	0.674	0.423	0.179	-0.155	1.503	1.593	0.111	*1.476	5.000	0.916
	2020	12	0.683	0.206	0.042	0.280	1.087	3.322	0.001			
	2019	4	1.058	0.449	0.202	0.177	1.938	2.353	0.019			
	2018	9	0.797	0.188	0.035	0.429	1.166	4.241	0.000			
	2017	3	0.464	0.355	0.126	-0.233	1.160	1.305	0.192			
	2015	2	0.538	0.431	0.186	-0.307	1.383	1.247	0.212			
Research method	Quantitative	12	0.779	0.200	0.040	0.387	1.172	3.893	0.000	*0.151	1.000	0.698
	Mixed	20	0.685	0.136	0.018	0.419	0.951	5.050	0.000			
Research design	Real-experimental pattern	2	0.801	0.813	0.661	-0.793	2.394	0.985	0.325	*1.444	3	0.695
	Semi-experimental pattern	17	0.857	0.159	0.025	0.546	1.168	5.397	0.000			
	Weak-experimental pattern	7	0.571	0.228	0.052	0.125	1.017	2.508	0.012			
	Non-experimental pattern	5	0.584	0.264	0.070	0.066	1.102	2.209	0.027			
Research model	Exp. control group	12	0.836	0.206	0.043	0.432	1.241	4.052	0.000	*0.607	2.000	0.738
	Single group	15	0.691	0.159	0.025	0.379	1.004	4.337	0.000			
	Correlational survey	5	0.584	0.267	0.071	0.061	1.107	2.190	0.029			
The educational level of the study group	5-8th grade	18	0.869	0.148	0.022	0.578	1.160	5.855	0.000	*6.856	3	0.077
	9-12th grade	7	0.884	0.270	0.073	0.355	1.413	3.274	0.001			
	Bachelor's degree	4	-0.022	0.326	0.106	-0.661	0.617	-0.067	0.947			
	Teachers	2	0.497	0.404	0.163	-0.295	1.288	1.229	0.219			
Sampling method	Not specified	3	0.428	0.305	0.093	-0.169	1.025	1.405	0.160	1.390	2.000	0.499
	Random	6	0.923	0.305	0.087	0.345	1.501	3.130	0.002			
	Non-random sampling methods	23	0.722	0.305	0.016	0.478	0.967	5.788	0.000			
Demographics of the study group	Not specified	3	1.248	0.421	0.177	0.424	2.073	2.967	0.003	*1.806	2.000	0.405
	Urban	23	0.658	0.132	0.017	0.399	0.917	4.978	0.000			
	Rural	5	0.747	0.325	0.106	0.110	1.384	2.300	0.021			
School type	Not specified	3	0.942	0.340	0.115	0.276	1.608	2.772	0.006	0.852	2.000	0.653
	Others	7	0.557	0.250	0.062	0.068	1.047	2.230	0.026			
	State school	22	0.727	0.138	0.019	0.457	0.997	5.274	0.000			
Research topic	Not specified	2	0.997	0.373	0.139	0.267	1.728	2.676	0.007	7.105	3.000	0.069
	Science	9	0.210	0.219	0.048	-0.219	0.638	0.959	0.338			
	Physics-chemistry-biology	6	0.919	0.228	0.052	0.472	1.367	4.029	0.000			
	STEM	15	0.808	0.149	0.022	0.516	1.100	5.421	0.000			
Duration of the application	Not specified	11	0.713	0.189	0.036	0.342	1.084	3.769	0.000	0.569	3	0.903
	1-3 weeks	7	0.588	0.245	0.060	0.108	1.069	2.398	0.016			
	4-7 weeks	8	0.862	0.273	0.074	0.328	1.397	3.161	0.002			
	8 or more weeks	6	0.746	0.275	0.075	0.208	1.284	2.717	0.007			
The person who made the application	Not specified	9	0.636	0.210	0.044	0.225	1.048	3.029	0.002	*3.042	2.000	0.218
	Exp. control researcher	6	0.323	0.287	0.082	-0.239	0.885	1.127	0.260			
	Exp. control teacher	17	0.871	0.154	0.024	0.570	1.172	5.670	0.000			
Computer usage	Not specified	6	0.552	0.238	0.057	0.086	1.018	2.321	0.020	2.977	2.000	0.226
	Used	18	0.641	0.151	0.023	0.344	0.938	4.236	0.000			
	Unused	8	1.085	0.245	0.060	0.604	1.566	4.420	0.000			
Coding usage	Not specified	6	0.552	0.235	0.055	0.091	1.012	2.348	0.019	2.857	2.000	0.240
	Used	17	0.649	0.148	0.022	0.359	0.940	4.379	0.000			
	Unused	9	1.083	0.249	0.062	0.594	1.571	4.346	0.000			
Robotics usage	Not specified	6	0.552	0.239	0.057	0.085	1.020	2.315	0.021	*1.393	2.000	0.498
	Used	9	0.611	0.215	0.046	0.191	1.032	2.848	0.004			
	Unused	17	0.850	0.162	0.026	0.533	1.167	5.255	0.000			
Algorithm usage	Not specified	7	0.551	0.231	0.053	0.099	1.003	2.387	0.017	*1.705	2.000	0.426
	Unused	25	0.792	0.129	0.017	0.540	1.044	6.151	0.000			
Flipped classroom	Not specified	7	0.551	0.231	0.054	0.097	1.005	2.381	0.017	1.568	2.000	0.457
	Unused	25	0.785	0.128	0.016	0.534	1.036	6.135	0.000			
Measurement tool type	Not specified	1	1.121	0.500	0.250	0.141	2.102	2.241	0.025	5.675	3.000	0.129
	Achievement test	8	1.165	0.238	0.056	0.700	1.631	4.905	0.000			
	Likert type	17	0.580	0.140	0.020	0.306	0.855	4.142	0.000			
	Other types of scale	6	0.559	0.221	0.049	0.125	0.992	2.528	0.011			
The person who develops measurement tools	Pre-existing tool	17	0.622	0.151	0.023	0.325	0.919	4.107	0.000	0.816	1.000	0.366
	Developed by the researcher	15	0.824	0.165	0.027	0.501	1.148	4.993	0.000			

Note. N: Number of primary studies; d: Effect sizes; SE: Standard error; & CI: Confidence intervals

APPENDIX C

Table C1. Heterogeneity analysis results for subgroups of moderators

Moderator	Subgroups	Heterogeneity			
		Q	df(Q)	p-value	I^2
Publication type	Article	60.830	2	0.000	67.121
	Other types of publications	80.444	10	0.000	87.569
	General	141.309	30	0.000	78.062
Language	Turkish	30.563	4	0.000	86.912
	English	94.466	26	0.000	72.477
	General	141.309	30	0.000	78.062
Country	Turkey	50.698	10	0.000	80.275
	USA	32.059	12	0.001	62.568
	Other countries	51.981	7	0.000	86.534
	General	141.309	31	0.000	78.062
Pilot study	Not specified	2.028	3	0.567	0.000
	Yes	88.472	12	0.000	86.436
	No	41.795	14	0.000	66.503
	General	141.309	31	0.000	78.062
Publication year	2021	2.024	1	0.155	50.603
	2020	86.725	11	0.000	87.316
	2019	1.285	3	0.733	0.000
	2018	31.554	8	0.000	74.647
	2017	1.624	2	0.444	0.000
	2015	1.427	1	0.232	29.911
	General	141.309	31	0.000	78.062
Research method	Quantitative	26.177	11	0.006	57.979
	Mixed	112.759	19	0.000	83.150
	General	141.309	31	0.000	78.062
Research design	Real experimental	0.025	1	0.875	0.000
	Semi-experimental	115.329	16	0.000	86.127
	Weak experimental	2.116	6	0.909	0.000
	Non-experimental	3.204	4	0.524	0.000
	General	137.962	30	0.000	78.255
Research model	Experimental-control group	93.415	11	0.000	88.225
	Single group	40.190	14	0.000	65.166
	Correlational survey	3.204	4	0.524	0.000
	General	141.309	31	0.000	78.062
The educational level of the study group	5-8th grade	58.26642	17	0.000	70.824
	9-12th grade	28.32269	6	0.000	78.816
	Bachelor's degree	32.30569	3	0.000	90.714
	Teachers	0.384347	1	0.535	0
	General	138.5565	30	0.000	78.34818
The sampling method of the study group	Not specified	0.062	2	0.969	0.000
	Random	12.485	5	0.029	59.951
	Non-random	104.689	22	0.000	78.985
	General	141.309	31	0.000	78.062
Demographics of the study group	Not specified	20.154	2	0.000	90.076
	Urban	116.161	22	0.000	81.061
	Rural	3.515	4	0.476	0.000
	General	141.103	30	0.000	78.739
School type of the study group	Not specified	20.788	2	0.000	90.379
	Others	2.745	6	0.840	0.000
	State school	109.502	21	0.000	80.822
	General	141.309	31	0.000	78.062
The research area of the application	Not specified	9.515	1	0.002	89.490
	Science	33.899	8	0.000	76.401
	Physics-chemistry-biology	28.052	5	0.000	82.176
	STEM	33.018	14	0.003	57.598
	General	141.309	31	0.000	78.062
Duration of the application	Not specified	39.394	10	0.000	74.616
	1-3 weeks	2.523	6	0.866	0.000
	4-7 weeks	23.958	7	0.001	70.782
	8 or more weeks	73.782	5	0.000	93.223
	General	141.309	31	0.000	78.062
The person who makes the application	Not specified	6.699	8	0.569	0.000
	Exp. control researcher	71.442	5	0.000	93.001
	Exp. control teacher	54.870	16	0.000	70.840
	General	141.309	31	0.000	78.062
Computer usage	Not specified	3.449	5	0.631	0.000
	Used	104.148	17	0.000	83.677
	Unused	25.354	7	0.001	72.391
	General	141.309	31	0.000	78.062
Coding usage	Not specified	3.449	5	0.631	0.000
	Used	106.260	16	0.000	84.943
	Unused	19.460	8	0.013	58.891
	General	141.309	31	0.000	78.062

Table C1 (Continued).

Moderator	Subgroups	Heterogeneity			
		Q	df(Q)	p-value	I^2
Robotics usage	Not specified	3.449	5	0.631	0.000
	Used	76.280	8	0.000	89.512
	Unused	53.398	16	0.000	70.036
	General	141.309	31	0.000	78.062
Algorithm usage	Not specified	3.449	6	0.631	0.000
	Unused	123.975	24	0.000	80.641
	General	141.309	31	0.000	78.062
Flipped classroom usage	Not specified	3.449	6	0.631	0.000
	Unused	127.548	24	0.000	81.184
	General	141.309	31	0.000	78.062
Measurement tool type	Achievement test	24.809	7	0.001	71.784
	Likert type	75.080	16	0.000	78.689
	Other types of scale	4.278	5	0.510	0.000
	General	141.309	31	0.000	78.062
The person who develops measurement tool	Pre-existing test	86.068	16	0.000	81.410
	Developed by the researcher	50.987	14	0.000	72.542
	General	141.309	31	0.000	78.062

◆ ⁂ ◆