Article

Examining the Effects of Artificial Intelligence on Elementary Students' Mathematics Achievement: A Meta-Analysis

Sunghwan Hwang [ID]

Department of Mathematics Education, Seoul National University of Education, Seoul 06637, Korea; ihwang413@gmail.com

Abstract: With the increasing attention to artificial intelligence (AI) in education, this study aims to examine the overall effectiveness of AI on elementary students' mathematics achievement using a meta-analysis method. A total of 21 empirical studies with 30 independent samples published between January 2000 and June 2022 were used in the study. The study findings revealed that AI had a small effect size on elementary students' mathematics achievement. The overall effect of AI was 0.351 under the random-effects model. The effect sizes of eight moderating variables, including three research characteristic variables (research type, research design, and sample size) and five opportunity-to-learn variables (mathematics learning topic, intervention duration, AI type, grade level, and organization), were examined. The findings of the study revealed that mathematics learning topic and grade level variables significantly moderate the effect of AI on mathematics achievement. However, the effects of other moderator variables were found to be not significant. This study also suggested practical and research implications based on the results.

1. Introduction

There is an increasing interest in artificial intelligence (AI) use in education to maximize students' learning outcomes [1–5]. Researchers have assumed that with the use of AI, barriers that hinder student learning (e.g., lack of qualified teachers and resources) could be eliminated, and the capacity of education could be maximized [2,6,7]. Increasing empirical studies have verified these assumptions and reported that AI has a positive effect on student learning outcomes [8–10]. In addition to the effect on student achievement, AI is important to ensure the sustainable development of our society. The United Nations Education Scientific and Cultural Organization (UNESCO) stated that sustainable development of our society could be achieved by ensuring the "inclusive and equitable quality of education and promote lifelong learning opportunities for all. AI technologies are used to ensure [them]" ([2], p. 12).

AI use in education is not an option; instead, it is a comparative educational movement. AI technology is widely integrated into our daily lives and many workforces (e.g., transportation, games, manufacturing, medical services, agriculture, and finance) to enhance the outcome and productivity of human work [11]. For example, as shown in Deep Mind's AlphaGo match with human Go players, the Go players are expected to learn and practice Go play with AI technology, as it could stimulate and facilitate human learning beyond traditional human-based learning [4]. Thus, one of the critical issues in the education and scientific research community is integrating AI into education [3,5,7,12]. Several governments, institutions, and industries have invested a lot of money to facilitate the integration of AI in education to support teaching and learning [13]. The World Bank estimated that the investment in AI use in education reached USD 1047 billion between 2008 and 2019 [14]. Additionally, many countries have revised their curriculum to integrate AI technology into school education [2,15].

Along with these educational changes, a literature review on AI in education has been widely implemented across various fields, including higher education [7], K-12 education [1], student assessment [16], robotics [17], data mining [18], and intelligent tutoring system (ITS) [19]. However, these studies examined the research trends of a certain topic without defining a single subject. Thus, we have limited information on previous studies on AI's effects on mathematics education. Moreover, few review studies focusing on mathematics education examined research design characteristics, such as author institutions and country, AI type used, target grade levels, and research methods [20]. Thus, there is little synthesized information on how the use of AI affects student mathematics achievement.

Mathematics learning outcomes affect not only students' success in school but also their college entrance, future careers, and social development [21,22]. Moses and Cobb [23] claimed that mathematics is a "gatekeeper course" (p. VII), and mathematics achievement is related to civil rights issues as students tend to have different opportunities based on their mathematics achievement. The OECD [22] and the United Nations [24] have also highlighted that students should be equipped with mathematical knowledge and competencies to adequately respond to a rapidly changing society for sustainable development. Therefore, further meta-analysis is needed to examine whether AI provides new mathematics learning opportunities [3,5,13]. Additionally, studies analyzing the effects of moderating variables on the relationship between them are required.

Ahmad et al. [3] pointed out that despite the increasing attention on AI in education, "the question as to how AI impacts education remains" (p. 5). This study aims to fill these gaps by synthesizing previous empirical studies on the effects of AI on student mathematics achievement using meta-analysis. Moreover, this study explored the effects of moderating variables, including research characteristic variables (e.g., research type and design, [5,25]) and opportunity-to-learn variables (e.g., mathematics learning topic and intervention duration, [26,27]). In particular, this meta-analysis examined elementary students' mathematics achievement, considering that previous meta-analysis on AI has focused on secondary and post-secondary students [5,7,28]. Given that mathematics achievement at elementary school is regarded as the foundation of future mathematics learning and career choices [29,30], the findings of this study could highlight the value of AI and provide guidance for using it for mathematics teaching and learning.

2. Literature Review

2.1. AI in Education

In 1955, McCarthy [31] first used the term AI in a research workshop and defined the AI problem as "making a machine behave in ways that would be called intelligent if a human were so behaving" (p. 12). Since then, many scholars have proposed various definitions of AI along with the development of AI technology [32,33]. While there is no consensus, scholars commonly agree that AI is not limited to the types of technology. Instead, AI relates to technologies, software, methods, and computer algorithms used to solve human-related problems [34]. According to Akgun and Greenhow [32], AI is a "technology that builds systems to think and act like humans with the ability to achieve goals" (p. 433). Similarly, Baker et al. [33] defined AI as "computers which perform cognitive tasks, usually associated with human minds, particularly learning and problem-solving" (p. 10).

Unlike traditional computer technologies, which provide a fixed sequence without considering the individual's needs and knowledge, AI interprets patterns of collected information (e.g., student understanding and errors) and makes reasonable decisions to provide the next tasks and maximize outcomes [35,36]. Additionally, based on a continuous learning and thinking process, AI evaluates prior strategies' outcomes and devises new ones. Thus, AI is likely to positively affect student achievement, creative thinking skills, and problem-solving abilities [5,11,33,37].

The positive effect of AI on mathematics learning outcomes could be explained by cognitive improvement and affective development. Because AI helps students develop a positive attitude toward mathematics and engagement in mathematics learning [37],

students are more likely to focus on mathematics learning and are eager to devote more time and effort. However, a few studies have reported the non-significant effects of AI on student achievement [38]. Because students need to manage their learning process as active learners with little teacher support, some students might not be able to focus on their learning and lose interest in using AI [3,17].

Among the various types of AI, the most widely used AI in education are ITS, adaptive learning system (ALS), and robotics [15]. In mathematics education, these types of AI have been widely adopted to improve the outcomes of mathematics teaching and learning [20,28,39,40]. ITS evaluates individual students' mathematical understanding and preferences and provides personalized feedback and instruction at their own pace [25,28]. Similar to ITS, ALS also provides individualized learning opportunities based on students' needs.

While ITS and ALS present course contents, evaluate student progress, and provide personalized feedback, teachers could engage in student learning passes in the ALS environment [7]. Teachers could check student learning progress using the information provided by ALS and suggest appropriate learning activities. For example, teachers could use the data collected by ALS and devise teaching strategies to support student learning [33,37]. Moreover, the tasks in ALS contained both conventional lesson-based tasks and game-based tasks [41,42]. However, ITS provides "customized instruction and immediate feedback without teacher intervention" ([43], p. 16). Thus, some researchers have suggested differentiating ALS from ITS [5,7,15,20].

Educational robotics allows students to explore diverse mathematical ideas by manipulating robotics. Because they provide interactive feedback, using robotics helps students develop cognitive thinking skills and reasoning abilities [10,39,44]. For example, Hoorn et al. [45] used robotics to facilitate students' understanding of multiplication and found improvement in their mathematics achievement. Similar findings were reported in studies with other graders and mathematical topics [16,39,44].

2.2. Review of Previous Meta-Analyses

Several meta-analyses have been conducted to examine factors that affect student achievement and generalize the results [5,46]. Zheng et al. [5] reviewed 24 articles and reported that AI had a large effect size on student achievement under the random-effects model (g = 0.812). Additionally, the moderating effects of schooling level, sample size, discipline, organization (individual or group), AI hardware (e.g., smartphones or tablet computers), and roles of AI (e.g., tutoring or policy-making advisor) were significant. It could be interpreted that the relationship between AI and learning achievement varied according to the degree of those moderating variables. For example, the effect size of AI could be different based on the schooling level. However, the moderating effects of learning methods (e.g., problem-based learning or project-based learning), research design (true or quasi-experimental design), research settings (e.g., laboratory or classroom), intervention duration, and AI types (e.g., ITS, ALS, or testing and diagnostic systems) were not significant.

Lin et al. [47] examined 21 articles with 23 independent samples and found a medium overall effect size under the random-effects model (g = 0.515). Similar to the findings of the study by Zheng et al. [5], they found significant moderating effects of schooling level and discipline and non-significant moderating effects of AI types. However, these two studies [5,47] did not focus on mathematics achievement; instead, they examined the learning achievement of all disciplines as an outcome variable.

To the best of our knowledge, no meta-analysis has examined the overall effects of AI on elementary students' mathematics achievement. Previous studies have examined the effects of a certain type of AI (e.g., ITS and robotics) on mathematics achievement [17,25] or the effects of computer technology on elementary students' mathematics achievement.

Regarding the relationship between ITS and mathematics achievement, Steenbergen-Hu and Cooper [25] examined 26 articles consisting of 31 independent samples focusing on

elementary to high school students. They found that most studies focused on secondary school students (n = 23), and only three examined elementary school students. The authors also reported that the overall mean effect of ITS use on mathematics achievement was not significant under the random-effects model and had a negligible effect under the fixed-effects model (g = 0.05). Similarly, Fang et al. [38] examined 15 studies with 24 independent samples and reported a non-significant effect of ITS on mathematics and statistics achievement.

However, other studies have different findings. Steenbergen-Hu and Cooper [28] replicated their meta-analysis using college students' data and reported a medium overall effect size (g = 0.59). Furthermore, Ma et al. [19] examined 35 individual samples of elementary to post-secondary students and found a positive association between ITS and mathematics and accounting learning achievement (g = 0.35).

For the analysis of moderating effects, Steenbergen-Hu and Cooper [25] reported that the effects of mathematics learning topic, schooling level, sample size, research design, report type (e.g., journal article or non-journal), measurement timing, and measurement types (e.g., standardized test or research created test) were not significant under the random-effects model. However, the effect of intervention duration was significant, and a shorter duration resulted in higher learning outcomes. Similarly, Fang et al. [38] highlighted the non-significant moderating effect of schooling levels, implementation strategies (supportive or principal instructional materials), and measurement type. They also found that the shorter periods of intervention duration (one semester) resulted in a larger effect size than longer intervention (one year).

Regarding the effect of robotics on K-12 students' creativity and problem-solving skills, Zhang and Zhu [17] reported a large effect size (g = 0.821). The authors noted significant moderating effects of schooling level and non-significant effects of gender and intervention duration. Similarly, Athanasiou et al. [48] analyzed the relationship between robotics use and learning outcomes and found a medium effect size (g = 0.70). Whereas the moderating effect of intervention duration was significant (studies of 1–6 months had a larger effect size than studies longer than six months), the effect of schooling level was not significant.

While not focusing on AI, Liao and Chen [49] examined the effects of technology use on elementary students' mathematics achievement. They reviewed 164 studies and found a small effect size (g = 0.372). Moreover, they found significant moderating effects of teaching devices and intervention duration. Similarly, Harskamp (2014) found a positive effect of computer technology on elementary students' mathematics achievement (Cohen's d = 0.48). Additionally, the effects were positive on all mathematics domains (i.e., number sense, operation, geometry, and measurement), and no significant differences between them were reported.

2.3. Analytical Framework

This study used the opportunity-to-learn (OTL) model to analyze the effect of moderator variables on the relationship between AI and mathematics achievement in elementary school students. OTL was proposed by Carroll [27] to explain different student learning outcomes. Researchers focused on instructional time and learning topics to analyze the degree of OTL [26,27]. However, with the increasing research on OTL, scholars have suggested other variables that affect student achievement.

Stevens [50] suggested four variables of OTL, including instructional delivery, content coverage, content emphasis, and content exposure. The instructional delivery variable analyzed teachers' instructional strategies, teaching methods, and tools. The content coverage variable was related to the curriculum and examined whether students learned a certain topic in school. The content emphasis variable concerned which contents and skills (e.g., lower- or higher-order thinking skills) were emphasized in classrooms. The content exposure variable examined the time allocated to students to study a subject. Similarly, Brewer and Stasz [51] proposed curriculum content, instructional strategies, and instructional resources as elements of OTL.

This study developed an analytical framework based on the OTL framework [26,50,51] (see Figure 1). According to the OTL variables, each study was examined across five dimensions, including mathematics learning topic (content coverage), intervention duration (content exposure), AI type (instructional resources), grade level (target students), and organization (instructional strategies). Additionally, research characteristics, which might affect the overall effect size, were examined based on the previous meta-analysis [5,17,28,38]. The variable included research type, research design, and sample size. The specific information of the analytical framework is presented in the Methods section.
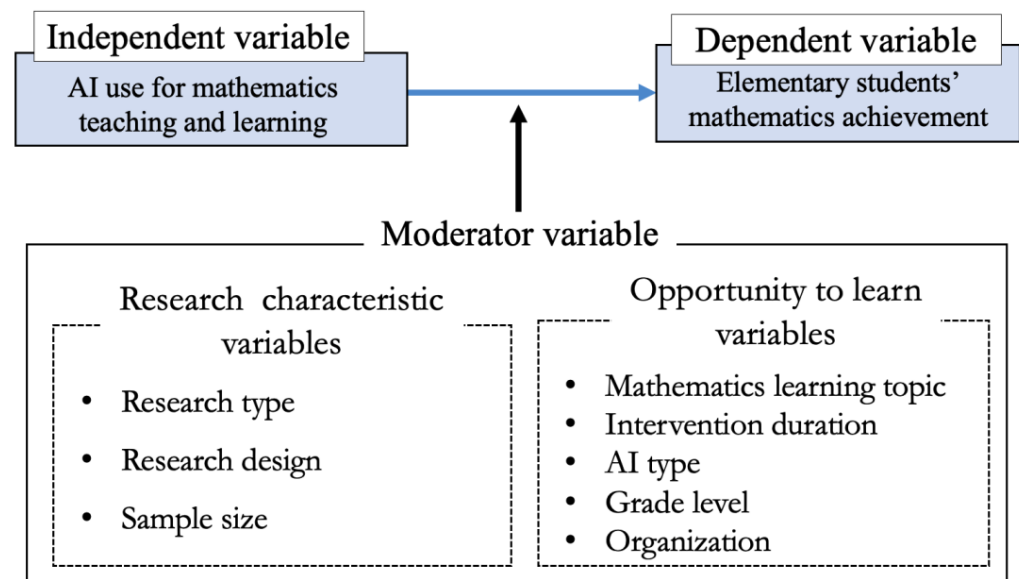
Figure 1. Analytical framework of this study.

2.4. The Present Study

The UNESCO and the OECD have highlighted the importance of using AI in education to ensure the inclusive and equitable quality of education, which leads to the sustainable development of our society [2,11]. Moreover, educators have suggested that AI use in education has a positive effect on student achievement [5,47]. Along with these arguments, researchers have conducted various meta-analysis studies to verify the assumption [5,17]. However, limited meta-analysis has been conducted to examine student mathematics achievement. Additionally, previous studies used to focus on secondary and post-secondary students [7,28,38]. Thus, we have limited information on the overall effect sizes of AI on elementary students' mathematics achievement and the roles of moderator variables affecting the relationship. Therefore, this study aims to fill the gaps in the literature by conducting a meta-analysis. The research questions of the study are as follows:

RQ 1. *Does AI use significantly improve elementary students' mathematics achievement?*

RQ 2. *Which variables moderate the overall effects of AI on elementary students' mathematics achievement?*

3. Methods

3.1. Article-Selection Process

Several selection criteria were used to retrieve relevant articles on June 2022. First, articles including AI [7,20], mathematics education, and elementary education-related terms (see Table 1) in titles or abstracts were collected using six databases (Web of Science, Education Source, ERIC, ScienceDirect, Taylor & Francis Online, and ProQuest Dissertation). Despite being one of the most widely utilized citation databases globally, Scopus was not used in this study as journals with a relatively short history tend to be excluded from the Scopus database due to its strict standards. Second, only English-written articles

were included, and non-English-written articles were excluded. Moreover, articles published before 2000 were excluded. AI in education has impressively grown and evolved since the beginning of the new millennium, along with the development of AI technology [52]. Since other systematic review studies tend to examine articles published after 2000 (e.g., [4,5]), the meaning of AI in those articles might differ from those published before 2000. Thus, articles published between January 2000 and June 2022 were included in this study. Third, all collected articles were imported into EndNote 20, and duplicated articles were excluded. One thousand twenty-four articles were obtained through these screening processes (see Figure 2).

Table 1. Search string used for retrieving relevant articles (keywords *).

AI-Related Terms	Mathematics-Education-Related Terms	Elementary-Education-Related Terms
"artificial intelligence" "deep learning" "machine learning" "chatbot" "robot"* "intelligent tutor"* "automated tutor"* "neural network"* "expert system" "intelligent system" "intelligent agent"* "virtual learning" "natural language processing"	"mathematics" "math" "geometry" "arithmetic" "addition" "subtraction" "multiplication" "division" "fraction" "decimal"	"elementary" "primary" "Grade 1" to "Grade 6" "first grade" to "sixth grade" "child"*
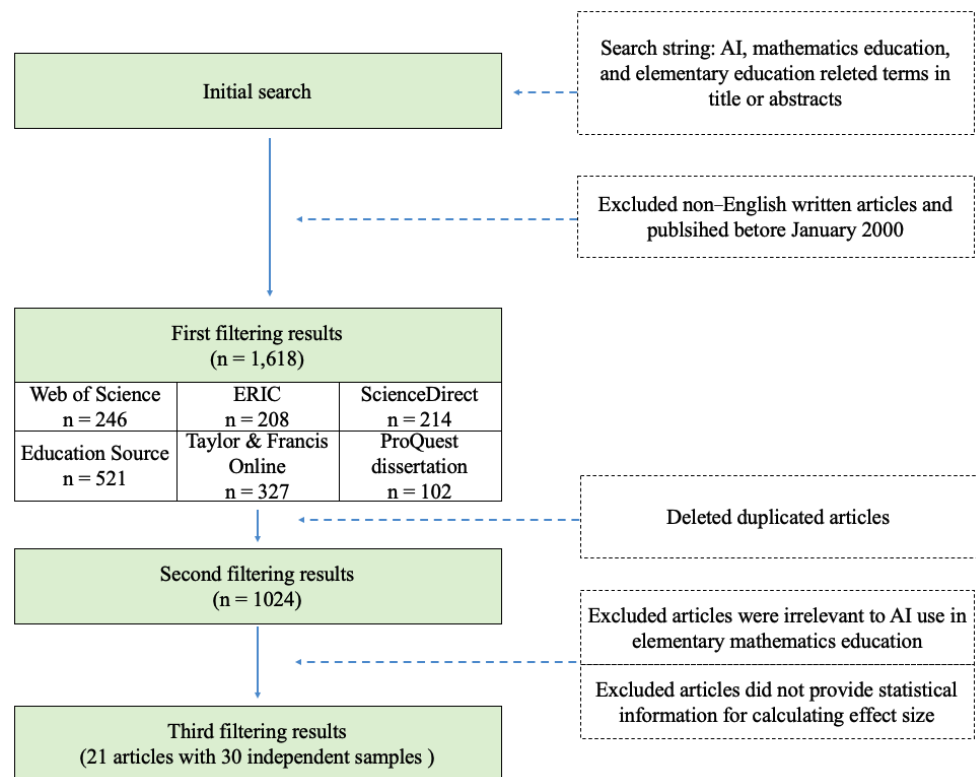
Figure 2. Article-selection process.

Fourth, I read each article and selected articles examining the effects of AI on elementary students' mathematics achievement and providing statistical information to calculate the effect size. For example, studies focused on science and engineering education and

analyzing secondary and post-secondary school students were excluded. Moreover, this study examined both experimental studies with the control group and non-experimental studies without the control group, following the guidance of a previous meta-analysis on education [53]. Because educational studies are likely to have a limited number of experimental studies, excluding non-experimental studies might not be able to provide sufficient data to implement meta-analysis [53]. A total of 21 articles with 30 independent samples were obtained through this screening process.

3.2. Coding Procedure

This study developed data-driven codes grounded in the collected data [54], but the previous literature (e.g., [5,50]) was reviewed to guide and check theoretical sensitivity [55]. Each moderator variable included several sub-categories considering the characteristics of collected articles (see Table 2). For example, this study only developed eight codes among the various mathematics learning topics because other topics did not exist in the collected articles. Additionally, while multiplication is a sub-area of arithmetic, this study categorized them into different codes, as some studies only examined multiplication [45], and other studies examined more than two arithmetic skills [56]. Moreover, the AI type was classified into ALS, ITS, and robotics, as other types of AI were not used. The validity of the developed codes and coding process was verified by three researchers who have expertise in mathematics education and educational science, while they did not participate in the study as co-authors.

Table 2. Coding schemes.

Dimension	Variable	Sub-Category
Research characteristics	Research type	-Journal paper, Dissertation
	Research design	-Experimental study, Non-experimental study
	Sample size	-1–40, 41–80, 81–120, Over 120
Opportunity to learn	Mathematics learning topic	-Determining areas, arithmetic, decimal numbers, finding patterns, fractions, multiplication, ratio and proportion, spatial reasoning
	Intervention duration	-1–5 h, 6–10 h, over 10 h
	AI type	-ALS, ITS, Robotics
	Grade level	-Grade 1 to 6 and mixed grade
	Organization	-Group work, Individual learning

3.3. Data Analysis

A comprehensive meta-analysis program [57] was used to examine the collected articles. First, Q statistics and I^2 values were calculated to examine heterogeneity. The Q statistics examined the common population effect size of all studies, and a significant Q value indicated that the effect sizes significantly vary across studies [57]. I^2 value examined the proportion of variability in the results across studies, which is not due to chance but real differences. If the I^2 value was greater than 50%, it would be appropriate to use the random-effects model and examine the reasons for the variance in effect sizes by implementing moderator analysis [46,57].

Second, the effect sizes of individual studies were calculated to estimate overall effects. When effect sizes were not reported in the study, relevant data (e.g., mean, standard deviation, sample size, and correlation) were used to calculate the effect size. This study used Hedges' g [58] to calculate the effect size because Cohen's d is likely to overestimate the outcomes [46]. The value of Hedges' g can be interpreted as small (between 0.20 and 0.50), medium (between 0.50 and 0.80), and large effects (higher than 0.80) [57]. When a

single study reported effect sizes of different samples (e.g., one for females and another for males), each effect size was treated as an individual sample study [53].

Third, the publication bias was examined to check whether articles with statistically significant results were more likely to be included in the meta-analysis than studies with non-significant results [57]. The funnel-plot method analyzed the distribution of the effect sizes with visualized information. A symmetrical distribution around the mean indicates an unbiased effect size, whereas an asymmetrical distribution represents a publication bias. Thus, when asymmetry distribution was detected, the trim-and-fill method was used to compare the difference between observed and adjusted (e.g., imputed) effect sizes. Moreover, classic fail-safe N and Orwin's fail-safe N tests were used to examine the publication bias [57].

Fourth, a moderator analysis was conducted to examine the effects of different variables on the overall effect sizes. The significant differences among groups were examined using a Q_B test, which is similar to ANOVA. Thus, a significant Q_B indicates heterogeneity among the groups and indicates that the effect sizes could be partially affected by the moderator variables [46].

4. Results

4.1. Overall Effect Size of AI on Mathematics Achievement

The review process identified 21 articles consisting of 30 independent samples. A total of 3 of the 30 independent samples had a negative effect, and 27 had a positive effect. The result of heterogeneity analysis showed a significant variance among effect sizes. The Q statistics was significant (Q = 83.225; *df* = 29; *p* < 0.001), and the I^2 value was moderately high (I^2 = 65.155). These results indicated that the variance among effects was partially affected by other variables and supported the necessity of moderator analysis. Considering the results of heterogeneity analysis and characteristics of this study (examining the overall effect sizes with different mathematics learning topics, AI type, and samples), this study adopted the random-effects model to analyze the collected data [57].

The overall mean effect size was small and significant. Under the random-effects model, the overall weighted mean effect size was 0.351 (SE = 0.072, k = 30, 95% CI: 0.221–0.471, Z = 5.756, *p* < 0.001). Regarding report type (see Table 3), 26 effect sizes were reported as published journal articles, and 4 effect sizes were reported as unpublished dissertations. The summary effects of journal articles (g = 0.368*) was higher than the summary effects of dissertations (g = 0.194). However, the difference was not significant (Q_B = 0.708, *df* = 1, *p* = 0.400).

Table 3. Moderator analysis for research characteristic variables.

Moderator Variable	Subgroup	K	Effect Size		95% CI		Between-Groups Effect
			g	SE	LL	UL	
Research type	Journal	26	0.368 *	0.065	0.242	0.495	Q_B = 0.708 *p* = 0.400
	Dissertation	4	0.194	0.197	−0.193	0.581	
Research design	Experimental study	16	0.284 *	0.086	0.115	0.453	Q_B = 1.253 *p* = 0.263
	Non-experimental study	14	0.422 *	0.088	0.249	0.594	
Sample size	1–40	9	0.545 *	0.130	0.290	0.801	Q_B = 2.858 *p* = 0.414
	41–80	12	0.290 *	0.099	0.096	0.484	
	81–120	6	0.288 *	0.140	0.013	0.562	
	More than 120	3	0.310	0.180	−0.043	0.662	

Note. LL and UL refer to lower and upper limits. * *p* < 0.05.

With regard to the research design, 16 studies used experimental conditions, while 14 studies used non-experimental conditions. However, the difference between the summary effects of experimental studies (g = 0.284 *) and the non-experimental studies

(g = 0.422 *) was not significant (Q_B = 1.253, df = 1, *p* = 0.263). Moreover, the sample size did not show a significant moderating effect (Q_B = 2.858, *df* = 3, *p* = 0.414). In summary, the non-significant moderating effects of three research characteristic variables indicated that the overall effect sizes were not significantly affected by the differences in research type, research design, and sample size. Therefore, this study combined articles with different research characteristics and analyzed them to examine OTL variables' influence. Table 4 shows information on the selected studies.

Table 4. Information on selected studies.

Study	g	SE	Research Characteristics		Opportunity-to-Learn Variables				
			Type and Design	Sample Size	Learning Topic	Duration (hours)	AI Type	Grade Level	Organization
Bush [40]	0.315	0.128	J-EX	297	Fractions	10	ALS	4, 5	Individual
Christopoulos et al. [59]	0.060	0.200	J-EX	100	Arithmetic	8	ALS	3	Individual
Chu et al. [60]	0.545	0.183	J-EX	124	Fractions	1	ITS	5	Individual
Chu et al. [42]	0.777	0.193	J-EX	116	Fractions	2	ALS	3	Individual
Fanchamps et al. [39]-(1)	0.194	0.187	J-Non	62	Finding patterns	9	Robotics	5, 6	Group
Fanchamps et al. [39]-(2)	0.09	0.174	J-Non	62	Finding patterns	9	Robotics	5, 6	Group
Francis et al. [44]	0.959	0.199	J-Non	37	Spatial reasoning	Over 30	Robotics	4	Group
González-Calero et al. [10]-(1)	0.118	0.242	J-EX	74	Spatial reasoning	2	Robotics	3	Group
González-Calero et al. [10]-(2)	0.636	0.249	J-EX	68	Spatial reasoning	2	Robotics	3	Group
Hoorn et al. [45]	0.83	0.134	J-Non	75	Multiplication	—	Robotics	—	Individual
Hou et al. [61]	0.554	0.223	J-Non	53	Decimal numbers	3	ALS	5, 6	Individual
Hwang et al. [12]-(1)	−0.013	0.192	J-EX	109	Determining areas	2	ALS	5	Individual
Hwang et al. [12]-(2)	0.418	0.194	J-EX	109	Determining areas	2	ALS	5	Individual
Julia and Antoli [62]	0.613	0.451	J-EX	21	Spatial reasoning	8	Robotics	6	Group
Laughlin [63]-(1)	−0.158	0.295	D-EX	46	–	–	Robotics	4	Group
Laughlin [63]-(2)	0.198	0.296	D-EX	46	–	–	Robotics	5	Group
Lindh and Holgersson [64]	0.114	0.110	J-EX	331	–	Over 50	Robotics	5	Group
Moltudal et al. [9]	0.577	0.171	J-Non	40	–	4	ALS	5, 6, 7	Individual
Ortiz [65]	0.394	0.369	D-EX	30	Ratio and proportion	15	Robotics	5	Group
Pai et al. [56]-(1)	0.303	0.213	J-EX	89	Arithmetic	3.5	ITS	5	Individual
Pai et al. [56]-(2)	0.169	0.212	J-EX	89	Arithmetic	3.5	ITS	5	Individual
Rau et al. [66]-(1)	0.216	0.128	J-Non	57	Fractions	5	ITS	4, 5	Individual
Rau et al. [66]-(2)	0.304	0.143	J-Non	57	Fractions	5	ITS	4, 5	Individual
Rau et al. [66]-(3)	0.136	0.138	J-Non	57	Fractions	5	ITS	4, 5	Individual
Rau et al. [67]-(1)	−0.182	0.178	J-Non	32	Fractions	–	ITS	3, 4, 5	Individual
Rau et al. [67]-(2)	0.169	0.166	J-Non	37	Fractions	–	ITS	3, 4, 5	Individual
Rittle-Johnson and Koedinger [8]-(1)	1.639	0.443	J-Non	13	Decimal numbers	2	ITS	6	Individual
Rittle-Johnson and Koedinger [8]-(2)	2.002	0.513	J-Non	13	Decimal numbers	2	ITS	6	Individual
Ruan [68]	0.354	0.245	D-Non	18	–	1	ITS	3, 4, 5	Individual
Vanbecelaere et al. [41]	0.185	0.243	J-EX	68	Arithmetic	3	ALS	1	Individual

Note. (1), (2), and (3) show different samples within the same study. The dashed line ('–') shows that the study did not provide information on the variable. Note: Regarding research type and design, J and D indicate published journal articles and unpublished dissertations, respectively. EX and Non represent experimental (with control group) and non-experimental (without control group) studies.

4.2. Publication Bias

The analysis of the funnel plot showed an almost symmetrical distribution, indicating the absence of publication bias [57]. The result of the trim-and-fill test also revealed a similar outcome. As Figure 3 illustrates, one study might be missed to the left of the mean: the observed and imputed effects were represented as open and filled circles. However, the values of the observed mean effect size (open diamond) and the adjusted mean effect size with the trim-and-fill method (filled diamond) were almost similar. The results of other tests also support the absence of publication bias. The classic fail-safe N was 724, indicating that nearly 724 studies were needed to nullify the mean effect size found in the study. Similarly, Orwin's fail-safe N was 164 at the level of 0.05 [19]. Therefore, while there is little possibility of publication bias, the publication bias did not pose a significant threat to study validity, and the major findings of this study were valid.
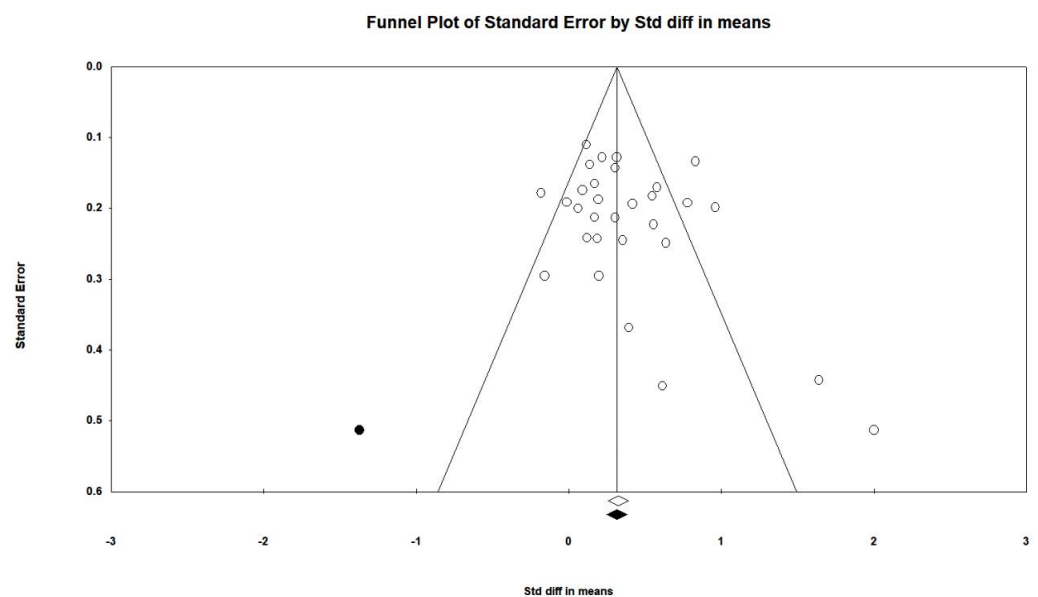
Figure 3. Funnel plot with trim-and-fill method.

4.3. Moderator Analysis

Based on the analytical framework of the study, the moderating effect of five OTL variables was examined. Studies that did not provide information on moderator variables were excluded during the analysis. Table 5 shows the results of the moderator analysis.

4.3.1. Mathematics Learning Topic

Regarding mathematics learning topics, fractions took the largest proportion (n = 8), followed by spatial reasoning (n = 4), arithmetic (n = 4), and decimal numbers (n = 3). The other topics included determining areas (n = 2), finding patterns (n = 2), multiplication (n = 1), and ratio and proportion (n = 1). AI use produced significant mean effect sizes in decimal numbers (g = 1.062 *), multiplication (g = 0.830 *), spatial reasoning (g = 0.600 *), and fractions (g = 0.276 *). However, the effect was not significant in determining areas, arithmetic, finding patterns, and ratio and proportion. The moderating effect was statistically significant (Q_B = 18.895, *df* = 7, *p* = 0.009), indicating that the effects of AI on mathematics achievement vary according to the mathematics learning topic.

4.3.2. Intervention Duration

Of the 25 samples, 17 took between one and five hours, 5 took between six and 10 h, and 3 took more than 10 h. It was found that the intervention duration of 1 to 5 h produced a larger effect size (g = 0.488 *) than that greater than 10 h (g = 0.463 *) and 6 and 10 h

(g = 0.210, $p > 0.05$). However, the differences across groups were not significant (Q_B = 2.330, *df* = 2, *p* = 0.408).

4.3.3. AI Type

The AI type was a principal technology used for mathematics teaching and learning. Eleven samples used robotics, and others used ITS (n = 11) and ALS (n = 8). The effectiveness of the AI was not significantly different across different types of AI (Q_B = 0.057, *df* = 2, *p* = 0.972). They all had positive effects on mathematics achievement. Robotics (g = 0.366 *) had the largest effect size, followed by ALS (g = 0.362 *) and ITS (g = 0.333 *).

Table 5. Moderator analysis for OTL variables.

Moderator Variable	Subgroup	K	Effect Size		95% CI		Between-Groups Effect
			g	SE	LL	UL	
Mathematics learning topic	Determining areas	2	0.201	0.204	−0.199	0.602	Q_B = 18.895 *p* = 0.009
	Arithmetic	4	0.177	0.179	−0.130	0.571	
	Decimal numbers	3	1.062 *	0.237	0.604	1.534	
	Finding patterns	2	0.140	0.199	−0.249	0.530	
	Fractions	8	0.276 *	0.094	0.092	0.461	
	Multiplication	1	0.830 *	0.254	0.333	1.327	
	Ratio and proportion	1	0.394	0.427	−0.443	1.231	
	Spatial reasoning	4	0.600 *	0.170	0.265	0.933	
Intervention duration	1–5 h	17	0.488 *	0.100	0.291	0.684	Q_B = 2.330 *p* = 0.408
	6–10 h	5	0.210	0.155	−0.093	0.514	
	Over 10 h	3	0.463 *	0.200	0.071	0.856	
AI type	ALS	8	0.362 *	0.117	0.133	0.591	Q_B = 0.057 *p* = 0.972
	ITS	11	0.333 *	0.110	0.130	0.536	
	Robotics	11	0.366 *	0.107	0.155	0.576	
Grade level	1st Grade	1	0.185	0.303	−0.408	0.779	Q_B = 16.688 *p* = 0.005
	3rd Grade	4	0.403 *	0.142	0.124	0.681	
	4th Grade	2	0.539 *	0.212	0.123	0.956	
	5th Grade	8	0.251 *	0.097	0.060	0.441	
	6th Grade	3	1.378 *	0.289	0.812	1.945	
	Mixed	11	0.240 *	0.074	0.094	0.386	
Organization	Group work	10	0.298 *	0.113	0.076	0.520	Q_B = 0.325 *p* = 0.569
	Individual learning	20	0.375 *	0.074	0.230	0.520	

Note. Because some studies did not provide information on moderator variables, the sum of some subgroups was not 30. LL and UL refer to lower and upper limits. * $p < 0.05$.

4.3.4. Grade Level

Of the 29 independent samples, almost one-third of the studies examined mixed grades (n = 11). Studies with fifth graders took the second largest proportion (n = 8), followed by third (n = 4) and sixth (n = 3) graders. There was no study examining second graders. Effect sizes were significantly different across the grade levels (Q_B = 16.688, *df* = 5 *p* = 0.005), indicating that the effects of AI on mathematics achievement were different across the grade levels. Except for studies on first graders, the effect sizes were statistically significant. The studies with sixth graders (1.378 *) had the largest effect, followed by fourth (0.539 *) and third (0.403 *) graders.

4.3.5. Organization

To measure the effectiveness of AI on mathematics achievement, 10 samples used the group work strategy, and 20 samples used the individual learning strategy. The effect sizes of both groups were significant: group work (g = 0.298 *) and individual learning (g = 0.375 *). The findings revealed no significant difference between the two groups (Q_B = 0.325, *df* = 1 *p* = 0.569).

5. Discussion

Recently, AI has been seen as an essential educational resource to maintain sustainable development and improve student achievement [2,11]. Despite the increasing attention to AI, relatively little is known about the overall effectiveness of AI on elementary students' mathematics achievement. Considering that mathematics achievement during the elementary school period affects students' future mathematics learning and career choices [23,29,30], this study examined the effects of AI on elementary students' mathematics achievement using a meta-analysis.

This study examined 21 articles with 30 independent samples regarding the first research question. This study found that the use of AI significantly and positively affected elementary students' mathematics achievement. The overall effect size was found to be 0.351 under the random-effects model. The positive effects of AI on mathematics achievement could be explained by responsive teaching [69] and constructionism [70]. According to the National Council of Teachers of Mathematics (NCTM, [71]), effective mathematics teaching should be built on responsive teaching strategies. Teachers should examine students' mathematical understanding and errors and adjust their instruction to meet student needs appropriately [72]. In responsive teaching environments, students' ideas are used as the basis for teacher instructions. Thus, students might engage in meaningful mathematical learning with a high motivation level, resulting in improved mathematical knowledge, skills, and achievement [69,71,73]. However, due to a lack of mathematical knowledge, poor teaching strategies, and large class sizes, some mathematics teachers find it challenging to adopt responsive teaching in classrooms [69].

AI could overcome such limitations. AI-based learning technologies, such as ALS and ITS, could provide personalized teaching and feedback on the basis of students' mathematical understanding. With advanced technology (e.g., machine learning), AI contains a variety of mathematical teaching strategies, problems, and information [1,2]. By analyzing students' problem-solving strategies and providing relevant mathematical knowledge and problems to increase student understanding through one-on-one interaction, AI could precisely examine and determine students' mathematical understanding [40]. Because AI can identify student progress and predict student achievement along with adaptive learning systems, it can provide efficient, responsive teaching environments for mathematics learning [5]. Moreover, unlike the traditional mathematics classroom, students can receive the same instructions whenever they want.

The positive effects of robotics on mathematics achievement are consistent with the theory of constructionism [70]. According to constructionists, developed based on Piaget's constructivism, learning is achieved through individuals' interaction with their environments [70,74]. Because individuals examine their ideas by using environments as a tool for learning, they can solidify previous knowledge and construct new knowledge. Similarly, the use of robotics in education could offer students a new learning environment. Through multiple iterations, students could develop computer programs and manipulate robotics. These meaningful experiences help them create new mathematical understanding and improve their problem-solving abilities and achievement [39].

However, such an effect is small compared to previous studies' effect sizes. Zheng et al. [5] found the average effect size of AI on student achievement was 0.812, and Lin et al. [47] reported an average effect size of 0.515. Moreover, the effect size was much smaller than the findings of Athanasiou et al. [48] (g = 0.70) and Zhang and Zhu [17] (g = 0.821), who examined the effects of robotics on student achievement.

Different target samples and disciplines might explain this difference. The current study only reviewed articles analyzing elementary students' mathematics achievement. However, the studies examined covered elementary students to adult workers across all disciplines. For example, Zheng et al. [5] reviewed 24 articles, including elementary students (n = 6), secondary students (n = 7), post-secondary students (n = 10), and working adults (n = 1). Moreover, the meta-analysis covered natural science (n = 6), social science (n = 9), and engineering and technology science (n = 9) studies. Similarly, Athanasiou

et al. [48], Lin et al. [47], and Zhang and Zhu [17] reviewed articles examining the effects of AI (or robotics) on K-12 students' learning achievement across all disciplines.

When comparing previous meta-analysis studies focusing on mathematics achievement, the effect size of this study (g = 0.351) was smaller than Steenbergen-Hu and Cooper [28] (g = 0.65), larger than Fang et al. [38] and Steenbergen-Hu and Cooper [25] (g was non-significant), and comparable with that of Ma et al. [19] (g = 0.35) and Liao and Chen [49] (g = 0.372). However, those studies examined the effectiveness of ITS [25,28,38] and technologies [49], not AI. The findings of this study distinguish the effectiveness of AI from other types of technologies in previous reviews and provide new insight into the roles of AI in elementary students' mathematics learning.

Regarding research question 2, this study conducted moderator analysis with research characteristics and OTL variables. The three research characteristic variables were found to be insignificant, including research type (journal paper or dissertation), research design (experimental or non-experimental study), and sample size. These findings revealed that the effectiveness of AI was not affected by those research characteristics. Steenbergen-Hu and Cooper [25] also reported that the moderating effects of research type, research design, and sample size were not significant in the relationship between ITS and students' mathematics achievement. Given similar findings in both studies, the non-significant effects of research characteristics might be common among studies examining student mathematics achievement with technology devices. However, further studies need to be conducted to examine this assumption.

With regard to the moderator analysis with five OTL variables, the effects of mathematics learning topic and grade level were significant. However, the moderating effects of intervention duration, AI type, and organization were not significant. The non-significant moderating effects indicated that AI positively affects elementary students' mathematics achievement, regardless of those variables. It is possible to infer from the significant moderating effects that the characteristics of the two moderating variables may affect the effects of AI to increase or decrease.

As for mathematics learning topics, previous studies have reported that the effects of AI on student achievement differ by discipline [17,47]. This study supported the findings of those studies and extended them by analyzing specific learning topics within mathematics. AI produced positive effects on decimal numbers, multiplication, spatial reasoning, and fractions, whereas the effects of AI on determining areas, arithmetic, finding patterns, and ratio and proportion were not significant.

The non-significant effect sizes of those mathematical topics might originate from the very small effect sizes of some studies. For example, in a study examining the effect of robotics on students' ability to find patterns, Fanchamps et al. [39] reported a small positive effect of robotics on student achievement (g < 0.20). Similarly, Pai et al. [55] analyzed the effects of ITS on fifth graders' arithmetic ability and reported that ITS was more effective than traditional teacher instruction with a very small effect size (g = 0.169). Christopoulos et al. [58] analyzed the effects of ALS on student achievement in determing areas and found almost no effect size (g = 0.060). Thus, it is possible to infer that while the effects of AI on those mathematical topics were not statistically significant, AI usually has a positive effect on students' learning. However, as a meta-analysis, this study was unable to fully explain the reasons for the small effect sizes of those topics. Further empirical studies can examine them.

However, these findings contradict the findings of Harskamp's (2014) study. Harskamp reported that there was no significant difference across mathematics learning topics and that computer technology positively affects all mathematics learning topics. Thus, it would be safe to assume that the effects of AI on mathematics achievement might differ from the effects of computer technology. This assumption is reasonable considering that some conventional computer technology is operated by teachers, whereas almost all AI (e.g., ITS, ALS, and robotics) tends to be used by students with autonomy, and teachers support their learning as facilitators [3,13].

This study also found that the effects of AI on elementary students' mathematics achievement have different effects according to different grade levels. Sixth-graders showed the largest effect size, followed by fourth- and third-graders. Moreover, the effect of AI on first-graders' mathematics achievement was not significant. Vanbecelaere et al. [41] examined the effects of ALS on first-graders' mathematic achievement in arithmetic and found a very small effect size (g = 0.185). This showed that even though using ALS was effective for students' mathematics learning, it did not significantly improve their achievement.

This may be due to the fact that first-graders are too young to learn mathematics without teachers' active guidance. When students learn mathematics with ALS and ITS, they need to manage their learning with learning strategies and skills [3,20,25]. Because first-graders are easily distracted by environmental factors, they might be unable to focus on mathematics learning. Consequently, due to their lack of learning strategies and attention to mathematics learning, the positive effects of AI on mathematics achievement might decrease. In contrast, students of other grades could develop new learning opportunities by exploring various mathematical ideas with personalized feedback [10,39,44]. Additionally, AI might help students develop learning motivation and a positive attitude toward mathematics [37], enhancing mathematics achievement.

The effect sizes of AI were not significantly different across three types of intervention duration. Interestingly, the shortest duration (between one to five hours) had a larger effect than the intervention duration of over 10 h. Additionally, the effect of intervention duration between six and 10 h was not significant. This result indicated that the longer intervention does not guarantee higher effects. Other studies have reported similar findings [28,38]. These results could be interpreted that when students use new AI technology for the first time, they might have curiosity and enthusiasm for using them, resulting in an improvement in mathematics achievement. However, their attention might dimmish, and the effects of AI might lower over time [17]. Then, the intervention with a duration greater than 10 h might reinforce their learning motivation, and the familiarity and adaptability to AI help them achieve high mathematics achievement [17].

The moderating effects of AI type (i.e., ALS, ITS, and robotics) and organization (i.e., group work or individual learning) were not significant, and the effects of all sub-groups were significant. These findings support the findings of previous studies examining the relationship between AI and achievement (e.g., [5]). Therefore, we could conclude that AI is an effective mathematics learning method for elementary students, regardless of AI type and organization.

6. Limitations

This study has five limitations. First, this study examined 21 studies with 30 independent samples focusing on elementary students' mathematics achievement. Thus, this study could not ensure the generalizability of the findings to other participants and contexts. Second, of the 30 independent samples, 14 samples did not use a control group and examined 1 group pre-post data. Although there was no significant difference between experimental and non-experimental studies, readers should be cautious when interpreting our findings. Third, due to the small amount of data performed at elementary school, this study could not compare the effectiveness of learning mathematics with AI and conventional teacher instruction. Fourth, this meta-analysis examined the effects of various moderator variables. However, some important variables affecting student mathematics achievements, such as teacher instructional style and classroom environment [29] and student socioeconomic status [30], were not controlled. If these variables had been included in the study, the findings of this study might be different. Fifth, this study only used six databases and collected English-written studies. Researchers who use other databases and include non-English-written studies might obtain different results. Considering the limitations of this study, future studies might use different grade levels, databases, and moderator variables to verify the study findings. Moreover, future studies might compare the effectiveness of mathematics learning with AI and conventional teacher instruction.

7. Conclusions and Implications

This study has three contributions to the existing literature. First, it examined the overall mean effect size of AI on mathematics achievement using a meta-analysis. This method helps synthesize the findings of previous empirical studies conducted at different places and times with different samples [5,46]. Thus, the findings of this study help researchers understand the current status of AI research in mathematics education and the effects of AI on mathematics achievement. Second, this study examined the effects of various moderating variables. This study considered three research characteristics and five OTL variables to examine the moderator effects. Thus, this study could provide more accurate information regarding how to implement AI to maximize its effectiveness. Third, this study focused on elementary students' mathematics achievement. Previous meta-analyses on AI focused on secondary [38] and post-secondary students [7]. Thus, we have limited information on how AI affects elementary students' mathematics achievement. This study extends previous studies on AI by examining elementary students.

This study suggested the following practical and research implications. As a practical implication, teachers need to implement AI for mathematics teaching and learning. This study shows that using AI positively affects elementary students' mathematics achievement. Thus, teachers should acquire relevant knowledge and skills and revise their lessons to integrate AI into their daily mathematics classrooms. School leaders should also support mathematics teachers to help them use AI with confidence. School leaders might provide professional development to help teachers understand how to operate AI and use it for teaching mathematics. Additionally, they could provide financial support to allow teachers to buy AI-related devices and tools.

However, teachers need to be cautious when using AI. Using moderator analysis, this study found that some environments are more effective in improving student mathematics achievement. For example, the intervention between six and 10 h was less effective than the intervention longer than 10 h. Moreover, AI was not effective in some mathematics learning topics (e.g., Determining areas) and first graders. Therefore, teachers should consider the effect of moderating variables and examine whether AI is effective in supporting student mathematics learning.

As a research implication, researchers need to report statistics information more accurately. This study used six databases to retrieve relevant articles. However, only 21 articles were used in this study. While many studies have examined AI to improve elementary students' mathematics achievement, most studies did not report accurate statistics that could be used to calculate the effect size. Accurate statistical information helps readers understand a study's findings and is useful for future studies.

Moreover, some studies did not use control groups and implemented a one-group pre–post research design. Considering that a one-group study only provides partial information on the data, using an experimental condition could provide more detailed information about the effectiveness of AI on elementary students' mathematics achievement.

Mathematics education researchers also need to implement more diverse studies to shed light on the effects of AI on student achievement. This study categorized AI types into ALS, ITS, and robotics, as other types of AI were not found. However, researchers have suggested different types of AI, such as teaching and diagnostic systems, assessment and monitoring systems, and agent systems [5]. For example, Kaoropthai et al. [75] developed a diagnostic test to examine students' reading abilities. Based on the test results, they classified students into different clusters and provided personalized tutoring using data mining techniques. Because the diagnostic system could analyze student strengths and weaknesses, it could successfully predict student performance and provide effective feedback. Thus, researchers studying elementary mathematics education need to use various types of AI to examine the effectiveness of AI on student achievement accurately.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zafari, M.; Bazargani, J.S.; Sadeghi-Niaraki, A.; Choi, S.-M. Artificial intelligence applications in K-12 education: A systematic literature review. *IEEE Access* **2022**, *10*, 61905–61921. [CrossRef]
2. Pedro, F.; Subosa, M.; Rivas, A.; Valverde, P. *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*; UNESCO: Paris, France, 2019.
3. Ahmad, S.F.; Rahmat, M.K.; Mubarik, M.S.; Alam, M.M.; Hyder, S.I. Artificial intelligence and its role in education. *Sustainability* **2021**, *13*, 12902. [CrossRef]
4. Paek, S.; Kim, N. Analysis of worldwide research trends on the impact of artificial intelligence in education. *Sustainability* **2021**, *13*, 7941. [CrossRef]
5. Zheng, L.; Niu, J.; Zhong, L.; Gyasi, J.F. The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interact. Learn. Environ.* **2021**, 1–15. [CrossRef]
6. Zhang, K.; Aslan, A.B. AI technologies for education: Recent research & future directions. *Compu. Edu.* **2021**, *2*, 100025. [CrossRef]
7. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *Int. J. Educ. Technol. High. Ed.* **2019**, *16*, 1–27. [CrossRef]
8. Rittle-Johnson, B.; Koedinger, K. Iterating between lessons on concepts and procedures can improve mathematics knowledge. *Brit. J. Educ. Psychol.* **2009**, *79*, 483–500. [CrossRef] [PubMed]
9. Moltudal, S.; Høydal, K.; Krumsvik, R.J. Glimpses into real-life introduction of adaptive learning technology: A mixed methods research approach to personalised pupil learning. *Design. Learn.* **2020**, *12*, 13–28. [CrossRef]
10. González-Calero, J.A.; Cózar, R.; Villena, R.; Merino, J.M. The development of mental rotation abilities through robotics-based instruction: An experience mediated by gender. *Brit. J. Educ. Technol.* **2019**, *50*, 3198–3213. [CrossRef]
11. OECD. *Artificial Intelligence in Society*; OECD Publishing: Paris, France, 2019.
12. Hwang, G.-J.; Sung, H.-Y.; Chang, S.-C.; Huang, X.-C. A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors. *Compu. Edu. Art. Intel.* **2020**, *1*, 100003. [CrossRef]
13. Chen, L.; Chen, P.; Lin, Z. Artificial intelligence in education: A review. *IEEE Access* **2020**, *8*, 75264–75278. [CrossRef]
14. Mou, X. Artificial intelligence: Investment trends and selected industry uses. *Int. Finance. Corp.* **2019**, *8*, 1–8.
15. Chen, X.; Xie, H.; Hwang, G.-J. A multi-perspective study on artificial intelligence in education: Grants, conferences, journals, software tools, institutions, and researchers. *Compu. Edu.* **2020**, *1*, 100005. [CrossRef]
16. González-Calatayud, V.; Prendes-Espinosa, P.; Roig-Vila, R. Artificial intelligence for student assessment: A systematic review. *Appl. Sci.* **2021**, *11*, 5467. [CrossRef]
17. Zhang, Y.; Zhu, Y. Effects of educational robotics on the creativity and problem-solving skills of K-12 students: A meta-analysis. *Edu. Stud.* **2022**, 1–19. [CrossRef]
18. Namoun, A.; Alshanqiti, A. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Appl. Sci.* **2020**, *11*, 237. [CrossRef]
19. Ma, W.; Adesope, O.O.; Nesbit, J.C.; Liu, Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. *J. Educ. Psychol.* **2014**, *106*, 901–918. [CrossRef]
20. Hwang, G.-J.; Tu, Y.-F. Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics* **2021**, *9*, 584. [CrossRef]
21. Gamoran, A.; Hannigan, E.C. Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educ. Eval. Policy.* **2000**, *22*, 241–254. [CrossRef]
22. OECD. *PISA 2018 Assessment and Analytical Framework*; OECD Publishing: Paris, France, 2019.
23. Moses, R.P.; Cobb, C.E. *Radical Equations: Math Literacy and Civil Rights*; Beacon Press: Boston, MA, USA, 2001.
24. UN. *Transforming Our World: The 2030 Agenda for Sustainable Development*; UN: New York, NY, USA, 2015.
25. Steenbergen-Hu, S.; Cooper, H. A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *J. Edu. Psyc.* **2013**, *105*, 970–987. [CrossRef]
26. OECD. *PISA 2012 Assessment and Analytical Framework*; OECD Publishing: Paris, France, 2013.
27. Carroll, J.B. A model of school learning. *Teach. Coll. Rec.* **1963**, *64*, 1–9. [CrossRef]
28. Steenbergen-Hu, S.; Cooper, H. A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *J. Edu. Psyc.* **2014**, *106*, 331–347. [CrossRef]
29. Little, C.W.; Lonigan, C.J.; Phillips, B.M. Differential patterns of growth in reading and math skills during elementary school. *J. Educ. Psychol.* **2021**, *113*, 462–476. [CrossRef] [PubMed]
30. Hascoët, M.; Giaconi, V.; Jamain, L. Family socioeconomic status and parental expectations affect mathematics achievement in a national sample of Chilean students. *Int. J. Behav. Dev.* **2021**, *45*, 122–132. [CrossRef]
31. McCarthy, J.; Minsky, M.L.; Rochester, N.; Shannon, C.E. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Mag.* **2006**, *27*, 12–14.

32. Akgun, S.; Greenhow, C. Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI Ethics* **2021**, *2*, 431–440. [CrossRef]

33. Baker, T.; Smith, L.; Anissa, N. *Educ-AI-Tion Rebooted? Exploring the Future of Artificial Intelligence in Schools and Colleges*; Nesta: London, UK, 2019.

34. Limna, P.; Jakwatanatham, S.; Siripipattanakul, S.; Kaewpuang, P.; Sriboonruang, P. A review of artificial intelligence (AI) in education during the digital era. *Adv. Know. Execu.* **2022**, *1*, 1–9.

35. Lameras, P.; Arnab, S. Power to the teachers: An exploratory review on artificial intelligence in education. *Information* **2021**, *13*, 14. [CrossRef]

36. Shabbir, J.; Anwer, T. Artificial intelligence and its role in near future. *J. Latex. Class.* **2018**, *14*, 1–10. [CrossRef]

37. Mohamed, M.Z.; Hidayat, R.; Suhaizi, N.N.; Mahmud, M.K.H.; Baharuddin, S.N. Artificial intelligence in mathematics education: A systematic literature review. *Int. Elect. J. Math. Edu.* **2022**, *17*, em0694. [CrossRef]

38. Fang, Y.; Ren, Z.; Hu, X.; Graesser, A.C. A meta-analysis of the effectiveness of ALEKS on learning. *Edu. Psyc.* **2019**, *39*, 1278–1292. [CrossRef]

39. Fanchamps, N.L.J.A.; Slangen, L.; Hennissen, P.; Specht, M. The influence of SRA programming on algorithmic thinking and self-efficacy using Lego robotics in two types of instruction. *Int. J. Technol. Des. Educ.* **2021**, *31*, 203–222. [CrossRef]

40. Bush, S.B. Software-based intervention with digital manipulatives to support student conceptual understandings of fractions. *Brit. J. Educ. Technol.* **2021**, *52*, 2299–2318. [CrossRef]

41. Vanbecelaere, S.; Cornillie, F.; Sasanguie, D.; Reynvoet, B.; Depaepe, F. The effectiveness of an adaptive digital educational game for the training of early numerical abilities in terms of cognitive, noncognitive and efficiency outcomes. *Brit. J. Educ. Technol.* **2021**, *52*, 112–124. [CrossRef]

42. Chu, H.-C.; Chen, J.-M.; Kuo, F.-R.; Yang, S.-M. Development of an adaptive game-based diagnostic and remedial learning system based on the concept-effect model for improving learning achievements in mathematics. *Edu. Technol. Soc.* **2021**, *24*, 36–53.

43. Crowley, K. *The Impact of Adaptive Learning on Mathematics Achievement*; New Jersey City University: Jersey City, NJ, USA, 2018.

44. Francis, K.; Rothschuh, S.; Poscente, D.; Davis, B. Malleability of spatial reasoning with short-term and long-term robotics interventions. *Technol. Know. Learn.* **2022**, *27*, 927–956. [CrossRef]

45. Hoorn, J.F.; Huang, I.S.; Konijn, E.A.; van Buuren, L. Robot tutoring of multiplication: Over one-third learning gain for most, learning loss for some. *Robotics* **2021**, *10*, 16. [CrossRef]

46. Hillmayr, D.; Ziernwald, L.; Reinhold, F.; Hofer, S.I.; Reiss, K.M. The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Compu. Edu.* **2020**, *153*, 103897. [CrossRef]

47. Lin, R.; Zhang, Q.; Xi, L.; Chu, J. Exploring the effectiveness and moderators of artificial intelligence in the classroom: A meta-analysis. In *Resilience and Future of Smart Learning*; Yang, J., Kinshuk, D.L., Tlili, A., Chang, M., Popescu, E., Burgos, D., Altınay, Z., Eds.; Springer: New York, NY, USA, 2022; pp. 61–66.

48. Athanasiou, L.; Mikropoulos, T.A.; Mavridis, D. Robotics interventions for improving educational outcomes: A meta-analysis. In *Technology and Innovation in Learning, Teaching and Education*; Tsitouridou, M., Diniz, J.A., Mikropoulos, T.A., Eds.; Springer: New York, NY, USA, 2019; pp. 91–102.

49. Liao, Y.-k.C.; Chen, Y.-H. Effects of integrating computer technology into mathematics instruction on elementary schoolers' academic achievement: A meta-analysis of one-hundred and sixty-four studies fromTaiwan. In Proceedings of the E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education, Las Vegas, NV, USA, 15 October 2018.

50. Stevens, F.I. The need to expand the opportunity to learn conceptual framework: Should students, parents, and school resources be included? In Proceedings of the Annual Meeting of the American Educational Research Association, New York, NY, USA, 8–12 April 1996.

51. Brewer, D.; Stasz, C. *Enhancing opportunity to Learn Measures in NCES Data*; RAND: Santa Monica, CA, USA, 1996.

52. Bozkurt, A.; Karadeniz, A.; Baneres, D.; Guerrero-Roldán, A.E.; Rodríguez, M.E. Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century. *Sustainability* **2021**, *13*, 800. [CrossRef]

53. Egert, F.; Fukkink, R.G.; Eckhardt, A.G. Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Rev. Educ. Res.* **2018**, *88*, 401–433. [CrossRef]

54. Grbich, C. *Qualitative Data Analysis: An Introduction*; Sage: London, UK, 2012.

55. Strauss, A.; Corbin, J. *Basics of Qualitative Research*; Sage Publisher: New York, NY, USA, 1990.

56. Pai, K.-C.; Kuo, B.-C.; Liao, C.-H.; Liu, Y.-M. An application of Chinese dialogue-based intelligent tutoring system in remedial instruction for mathematics learning. *Edu. Psyc.* **2021**, *41*, 137–152. [CrossRef]

57. Borenstein, M.; Hedges, L.V.; Higgins, J.P.; Rothstein, H.R. *Introduction to Meta-Analysis*; John Wiley & Sons: New York, NY, USA, 2021.

58. Hedges, L.V. Distribution theory for Glass's estimator of effect size and related estimators. *J. Educ. Stat.* **1981**, *6*, 107–128. [CrossRef]

59. Christopoulos, A.; Kajasilta, H.; Salakoski, T.; Laakso, M.-J. Limits and virtues of educational technology in elementary school mathematics. *J. Edu. Technol.* **2020**, *49*, 59–81. [CrossRef]

60. Chu, Y.-S.; Yang, H.-C.; Tseng, S.-S.; Yang, C.-C. Implementation of a model-tracing-based learning diagnosis system to promote elementary students' learning in mathematics. *Edu. Technol. Soc.* **2014**, *17*, 347–357. [CrossRef]

61. Hou, X.; Nguyen, H.A.; Richey, J.E.; Harpstead, E.; Hammer, J.; McLaren, B.M. Assessing the effects of open models of learning and enjoyment in a digital learning game. *Int. J. Artif. Intel. Edu.* **2022**, *32*, 120–150. [CrossRef]
62. Julià, C.; Antolí, J. Spatial ability learning through educational robotics. *Int. J. Technol. Des. Educ.* **2016**, *26*, 185–203. [CrossRef]
63. Laughlin, S.R. Robotics: Assessing Its Role in Improving Mathematics Skills for Grades 4 to 5. Doctoral Dissertation, Capella University, Minneapolis, MN, USA, 2013.
64. Lindh, J.; Holgersson, T. Does lego training stimulate pupils' ability to solve logical problems? *Compu. Educ.* **2007**, *49*, 1097–1111. [CrossRef]
65. Ortiz, A.M. *Fifth Grade Students' Understanding of Ratio and Proportion in an Engineering Robotics Program*; Tufts University: Medford, MA, USA, 2010.
66. Rau, M.A.; Aleven, V.; Rummel, N.; Pardos, Z. How should intelligent tutoring systems sequence multiple graphical representations of fractions? A multi-methods study. *Int. J. Artif. Intel. Edu.* **2014**, *24*, 125–161. [CrossRef]
67. Rau, M.A.; Aleven, V.; Rummel, N. Making connections among multiple graphical representations of fractions: Sense-making competencies enhance perceptual fluency, but not vice versa. *Instr. Sci.* **2017**, *45*, 331–357. [CrossRef]
68. Ruan, S.S. *Smart Tutoring through Conversational Interfaces*; Stanford University: Stanford, CA, USA, 2021.
69. Lampert, M. *Teaching Problems and the Problems of Teaching*; Yale University Press: New Haven, CT, USA, 2001.
70. Papert, S. What's the big idea? Toward a pedagogy of idea power. *IBM. Syst. J.* **2000**, *39*, 720–729. [CrossRef]
71. NCTM. *Principles to Actions: Ensuring Mathematical Success for All*; NCTM: Reston, VA, USA, 2014.
72. Dyer, E.B.; Sherin, M.G. Instructional reasoning about interpretations of student thinking that supports responsive teaching in secondary mathematics. *ZDM* **2016**, *48*, 69–82. [CrossRef]
73. Stockero, S.L.; Van Zoest, L.R.; Freeburn, B.; Peterson, B.E.; Leatham, K.R. Teachers' responses to instances of student mathematical thinking with varied potential to support student learning. *Math. Ed. Res. J.* **2020**, *34*, 165–187. [CrossRef]
74. Noss, R.; Clayson, J. Reconstructing constructionism. *Constr. Found.* **2015**, *10*, 285–288.
75. Kaaropthai, C.; Natakuatoong, O.; Cooharojananone, N. An intelligent diagnostic framework: A scaffolding tool to resolve academic reading problems of Thai first-year university students. *Compu. Edu.* **2019**, *128*, 132–144. [CrossRef]