

Technology Integration in Mathematics Education: Examining the Quality of Meta-Analytic Research

Jamaal Rashad Young

University of North Texas, 1155 Union Cir, Denton, TX 76203, USA

Email: jamaal.young@unt.edu

Abstrak

Penelitian ini bertujuan untuk mengevaluasi kualitas metodologi meta-analisis dalam penelitian pendidikan matematika. Pencarian komputerisasi dilakukan di ERIC, PsycINFO, ProQuest, dan JSTOR selama Januari 2016 yang berkaitan dengan meta-analisis untuk integrasi teknologi di kelas matematika. Identifikasi penelitian menggunakan istilah-istilah dalam materi matematika seperti *systematic review*, *literature review*, dan *meta-analysis*. Seluruh penelitian yang telah dikumpulkan, disaring, diekstraksi, dan kualitas metodologi dievaluasi dengan menggunakan 11 indikator *Revised Assessment of Multiple Systematic Reviews* (R-AMSTAR). Karakteristik meta-analisis yang terkait dengan kualitas pelaporan metodologi telah diidentifikasi. Hasil penelitian menunjukkan bahwa rata-rata kualitas pelaporan metodologi adalah baik, tetapi ada beberapa bidang yang menjadi perhatian. Implikasinya dilaporkan untuk meningkatkan penelitian dan praktek di kelas penelitian pendidikan matematika.

Kata Kunci: meta-analisis, kajian sistematis, integrasi teknologi, Res

Abstract

The aim of this study was to evaluate the methodological quality of meta-analyses in mathematics education research. Computerized searches were conducted in ERIC, PsycINFO, ProQuest, and JSTOR through January 2016 for meta-analyses pertinent to technology integration in the mathematics classroom. To identify studies, mathematics subject matter terms such as 'systematic review', 'literature review' and 'meta-analysis' were used. Retrieved studies were screened, extracted, and the methodological quality was evaluated using the 11-item Revised Assessment of Multiple Systematic Reviews (R-AMSTAR). The characteristics of meta-analyses associated with the methodological reporting quality were identified. Results suggest that on average the quality of methodological reporting is good, but there are several areas of concern. Implications are provided to improve research and classroom practice in this area.

Keywords: meta-analysis, systematic review, technology integration, Res

How to Cite: Young, J.R. (2017). Technology integration in mathematics education: Examining the quality of meta-analytic research. *International Journal on Emerging Mathematics Education*, 1(1), 71-86. <http://dx.doi.org/10.12928/ijeme.v1i1.5713>.

INTRODUCTION

Technological tools can enhance the delivery, clarity, and precision of mathematics instruction. However, the use of technological tools in the mathematics classroom must be strategic. The strategic use of technology in the mathematics classroom should be consistent with current research in order to support effective teaching (National Council of Teachers of Mathematics [NCTM], 2008). Based on the trends research examining the effective use of technology in the mathematics classroom has grown exponentially. To manage and inform this growth, critical syntheses of these results are ideal. The critical synthesis of these results should be guided by meta-analytic thinking. Meta-analytic thinking is defined as the prospective

formulation of study expectations and design by explicitly invoking prior effect sizes and the retrospective interpretation of new results by direct comparison with prior effects in the related literature (Thompson, 2002, p. 28). Numerous meta-analytic studies have investigated technology's effect on mathematics achievement and the factors that mediate these effects over the last three decades (Chan & Leung, 2014; Hembree & Dessart, 1986; Li & Ma, 2010). This growing body of research currently informs classroom practices in the U.S. and beyond. However, until researchers compare these effects across the literature, the reliability of these results across context is unknown.

Incorporating technology in the classroom is an important part of student success in mathematics (NCTM, 2000; Young & Young, 2012). High-quality literature summaries of research in mathematics education can inform scientific and instructional praxis. Thus, it is important that researchers and teachers have comprehensive research findings to inform their practice. Systematic reviews and meta-analysis are important empirical tools for the development of cumulative knowledge in social sciences. A meta-analysis is defined as the calculation of a summary estimate of treatment effect by pooling the results of multiple studies (Schmucker et al., 2013). The utility of meta-analysis is vast and multidimensional. Meta-analyses inform and direct future research and theoretical considerations (Cooper & Hedges, 2009; Hedges & Olkin, 2014). Meta-analysis supports the accumulation of evidence across studies (Cummings, 2013). The accumulation of similar and divergent findings promotes replication and validation. Meta-analysis also helps to guide emerging research by allowing scholars to identify lingering questions or interactions that require further analysis. As meta-analysis methodology came into fruition in mathematics education, studies examining the effects of technology-enhanced instruction on mathematics achievement grew as well.

Meta-analyses dedicated to technology integration in the mathematics classroom have increased considerably. A substantial number of meta-analysis provides uniquely nuanced summaries of these effects. Multiple operational definitions of technology integration exist, and substantial variability of foci is present across earlier studies. Yet, a comprehensive summary offering substantive and explicit implications for general technology enhanced mathematics instruction remains elusive. To inform technology integration in the mathematics classroom researchers must clearly define theoretical constructs through empirical specification. This has the potential to guide classroom applications with technology. Effect size reporting and meta-analytic thinking are considered major concerns in effective mathematics education research consumption and reporting. These reporting practices are indicators of methodological quality, clarity, and transparency.

The American Psychological Association (APA, 2010) and the American Educational Research Association (AERA, 2006) regularly advocate for the reporting of effect sizes and, more recently, considered meta-analytic thinking an extension to previous reporting practices. Numerous mathematics education scholars cite the benefits of effect size reporting (Capraro, 2004; Zientek, Capraro, & Capraro, 2008; Young, Young, & Hamilton, 2013). Effect sizes and confidence intervals are important elements of meta-analytic research and represent metrics for assessment and summarization of effects across studies (Young, 2016). Therefore, reviewing the reporting trends and examining the design quality in previous meta-analytic research is vital to the reliability, validity, and generalizability of technology integration research in the mathematics classroom.

Importance of Meta-Analysis Study Design and Reporting

A meta-analytic lens is a suitable empirical tool to categorize the best technology integration practices in mathematics education. Study quality is an important consideration when interpreting the results of research. Thus, as the results of multiple studies are aggregated

in meta-analytic research, it is increasingly more important to consider the design and reporting practices implemented. Meta-analysis is a research synthesis tool that uses summaries of effect sizes to generate empirical conclusions from ostensibly similar studies (Young, Ortiz, & Young, 2017). Because meta-analyses combine results across multiple studies, it is imperative that the design quality of included studies is assessed.

Meta-analysis involves: (1) summarizing several studies in terms of an effect size, and (2) combining the results to make summative inferences (Cooper, 2016). This process is traditionally completed in three steps: calculate the average effect size, test homogeneity, and detect moderators explaining any heterogeneity (Sanchez-Meca & Martin-Martinez, 1998). Maintaining the fidelity of these processes is essential to the validity and reliability of the meta-analysis results. The overall or mean effect size represents a key element of the meta-analytic review thus the calculation and reporting of these data must be undertaken with the utmost transparency (Cooper & Hedges, 2009). Researchers recommend that the details of the literature are rigorous and explicitly reported to avoid biased or erroneous results (Kepes et al., 2012; Rothstein, 2012). It is also recommended that any formulas or software used to complete the effect size calculations are referenced as well to increase the studies transparency.

The assessment of homogeneity in meta-analysis is also an important consideration in reporting and methodological quality. Two statistics are typically reported to assess the homogeneity in meta-analyses. The homogeneity test Q and the I^2 index, assess the true homogeneity of effect sizes (Borenstein et al., 2009). These statistics should be reported together because they are complementary, as the Q indicates the presents of homogeneity, while the I^2 quantifies the magnitude. Finally, arguably the most important step to increase the quality and rigor of a meta-analysis is to detect moderators and describe their influence on the heterogeneity. The detection of moderators is the key feature of any met-analytic study; because differences in strength and direction in effect sizes is identified here.

Rosenthal (1991) expounds, "The search for moderators is not only an exciting intellectual enterprise but indeed...it is the very heart of scientific enterprise" (p. 447). Moderators offer conditions for the effects that are theorized, thus informing researchers of the conditions in which the effects under investigation can be reliable (Schmidt & Hunter, 2014). This is the information vital to successful implementation of technology in the mathematics classroom across instructional contexts. Moderators are also important because they identify statistical interactions, which do not imply causation but rather add context to effect size results (Cooper & Patall, 2009). Given the distinctions in the associations moderators identify they are consistently placed in three categories. Moderators are categorized as either: (1) methodological variations, (2) theoretical constructs, or (3) study characteristics (DeCoster, 2004). Moderators are recognized for their ability to enhance theory development and increase the general richness of empirical work (Aguinis, Gottfredson, & Wright, 2011). Given the empirical merit of meta-analytic research and the contextualization offered by moderator analysis, examining the reporting and design quality across studies is practically and scientifically necessary.

Problem Statement

Because the methodological rigor employed in meta-analysis continues to increase, it is important to examine the affect of methodological rigor on effect size magnitude (Cooper, 2016; Moher, et al, 2015). Due to the high level of evidence meta-analyses are expected to provide, the importance of the methodological and reporting quality of these studies is an ongoing concern. In response, the American Psychological Association (APA) issued their Meta-Analysis Reporting Standards (MARS) as a means to share common meta-analytic practices across disciplines and to account for discipline specific priorities with psychology related disciplines such as education (APA, 2010). The MARS adheres to the assumption that methodological

qualities of meta-analysis are nuanced by discipline, thus particular aspects are more critical to one discipline compared to another (Kepes, McDaniel, Brannick & Banks, 2013). The MARS suggest that a common structure is followed, while allowing for some flexibility, which inhibits its use as a strict measure of reporting quality. Furthermore, within the psychological and social sciences common rubrics or measures of meta-analysis quality are underdeveloped. Yet, given the rich history of meta-analytic research in the medical sciences valid and reliable tools are more abundant. To address this problem, the current study synthesized ostensibly similar prior meta-analytic research and assessed the quality using a standardized assessment instrument adapted from the medical sciences.

Because valid and reliable tools to assess the quality of meta-analysis are currently lacking in the psychological and social sciences, study quality was assessed on an adapted version of the Revised Assessment of Multiple Systematic Reviews (R-AMSTAR). The R-AMSTAR is an instrument used to assess the methodological quality of systematic reviews and meta-analysis. Because the methodological rigor of prior meta-analysis varies across studies, the R-AMSTAR was used to assess the variation in the quality of included studies. For example, one concern with prior research is the inclusion of unstandardized outcome measures in prior meta-analysis and the possible inflation of overall effect size statistics. Using the R-AMSTAR to assess this and other methodological quality issues is important because the influence of problematic measures can be evaluated through moderator analysis to examine effects sizes across studies and inform future practice.

This critical synthesis was designed to inform research practice and consumption in mathematics education. Thus, this literature survey examined trends in results from meta-analytic research to provide recommendations for future research and instructional praxis. The results of this study summarize the overall quality of meta-analytic study designs across multiple settings and implementations. These results are important because they promote the expansion of meta-analytic thinking in mathematics education research.

RESEARCH METHOD

Comprehensive literature searches for articles written between 1980 and 2015 were conducted to produce the primary list of meta-analyses. Iterative electronic searches using educational databases identified and utilized from previous studies were conducted (JSTOR, ERIC, EBSCO, PsycINFO, and ProQuest). A combination of Boolean operators and the keywords (meta-analysis, research synthesis, literature review, literature synthesis, mathematics, achievement, technology, instructional technology, information communication technology, calculators, and computer-assisted instruction) were used to generate an initial pool of studies. Citations from retrieved studies were searched manually to identify any potentially missing meta-analysis. These methods identified 42 potential studies for preliminary review. The study titles and abstracts were read as part of the initial screening process and then the inclusion criteria presented below was applied.

The following inclusion criteria guided this literature survey:

1. The study examined the effects of digital technology applications, including computer-assisted instruction, integrated tutoring systems, technology based programs, or technological tools to improve mathematics achievement.
2. The studies used meta-analytic methods to calculate mean effect sizes.
3. The studies involved students in K-12 or post-secondary mathematics classrooms.
4. The primary dependent measure included quantitative measures of mathematics performance such as standardized test, researcher made test, or teacher-made test.

- Only studies conducted between 1980 and 2015 were included. The year 1980 was identified as a starting point given the significant influx of instructional technology that took place in the 80's.

Figure 1 presents the complete study inclusion and exclusion process. The initial pool of 42 studies was systematically screened and reduced to a final pool of 18. The majority of studies were removed for lack of effect size reporting and the absence of a digital technology focus. Pertinent data related to the research questions were extracted from the remaining studies.

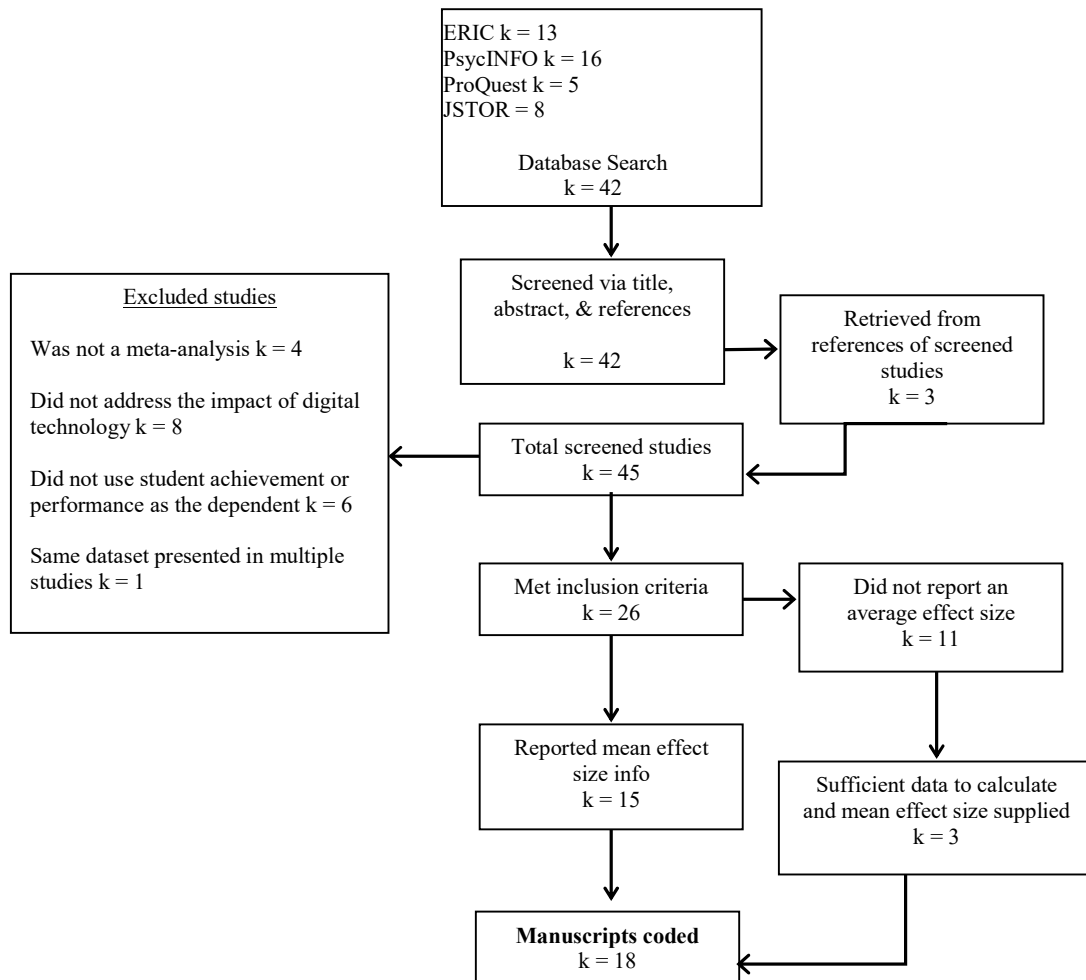


Figure 1. Study Inclusion Flowchart

The methodological quality of the meta-analyses was assessed on the R-AMSTAR tool, which is a reliable and valid instrument based on prior confirmatory factor analysis. The R-AMSTAR rating scale consists of 11 items scored on 3 to 5 criteria. Scores for each item are based on the inclusion of the related criteria and range from 0 to 4 points. The highest possible score on the instrument is 44. Sample items include, "Were the methods used to combine the findings of studies appropriate?" and "Was the likelihood of publication bias assessed?". Because scores are derived from individual items on an evaluative rubric the R-AMSTAR also allows researchers to

isolate specific design strengths and weaknesses across studies. After applying the R-AMSTAR the resulting inter-rater agreement was 86.3% (Cohen's $\kappa = .813$). To reconcile any discrepancies in the R-AMSTAR ratings, the two independent coders met to establish a consensus. Scores were then categorized based on the established R-AMSTAR grading scale: D = 0 to 11, C = 12 to 22, B = 23 to 33, and A = 34 to 44. Coding reliability was assessed by comparing the independent coding results obtained by two separate researchers from a random sample of ten studies. The initial inter-rater consistency was 95% and the researchers met to resolve the remaining inconsistencies in the coding results. The complete R-AMSTAR instrument is presented in Table 1.

Descriptive statistics including mean and standard deviations (SD) for continuous variables, and as frequencies (percentages) for categorical variables are presented. The frequency of each R-AMSTAR item was calculated. Mean difference effect sizes were calculated for between study characteristics (publication status, technology intervention, and number of effect size ≤ 20 or less, less than 100, and more than 100). Finally, 95% confidence intervals for the mean R-AMSTAR score for each categorical variable were plotted for comparison across groups.

Table 1. Adapted Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) items

| Statement | Score |
|--|---------------------------------------|
| 1. Was an 'a priori' design provided? | If it satisfies 3 of the criteria = 4 |
| a. 'a priori design' | |
| b. 'statement of inclusion criteria' | |
| c. research question includes (population, intervention/treatment/independent variable, comparison, prediction, outcome) | If it satisfies 2 of the criteria = 3 |
| | If it satisfies 1 of the criteria = 2 |
| | If it satisfies 0 of the criteria = 1 |
| 2. Was there duplicate study and data extraction? | If it satisfies 3 of the criteria = 4 |
| a. There should be at least two independent data extractors/coders as stated or implied. | |
| b. Statement of recognition or awareness of consensus procedure for disagreement. | If it satisfies 2 of the criteria = 3 |
| c. Disagreement among extractors/coders resolved properly as stated or implied. | If it satisfies 1 of the criteria = 2 |
| | If it satisfies 0 of the criteria = 1 |
| 3. Was a comprehensive literature search performed? | If it satisfies 4 of the criteria = 4 |
| a. At least two electronic searches should be searched. | If it satisfies 3 of the criteria = 3 |
| b. The report should include years and databases used (e.g. ERIC, PsychINFO). | |
| c. Keywords and terms must be stated AND | If it satisfies 2 of |

- where feasible the search strategy outline should be provided such that one can trace the filtering process of included articles.
- d. Journals were “hand-searched” or “manual searched”
- the criteria = 2
If it satisfies 1 or 0 of the criteria = 1
- 4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?**
- a. The authors should state that they searched for reports regardless of their publication type.
- b. The author should state whether or not they excluded any reports, based on their publication status, language etc.
- c. “Non-English papers were translated” or readers sufficiently trained in foreign language.
- d. No language restriction or recognition of non-English articles.
- If it satisfies 3 of the criteria = 4
If it satisfies 2 of the criteria = 3
If it satisfies 1 of the criteria = 2
If it satisfies 0 of the criteria = 1
- 5. Was a list of studies (included and excluded) provided?**
- a. Table/list of figure if included studies, a reference list does not suffice.
- b. Table/list/figure of **excluded** studies either in the article or in supplemental source. (Excluded studies refers to those studies seriously considered on the basis of the title and/or abstract, but rejected after reading the body of the text).
- c. Author satisfactorily/sufficiently stated the reason for exclusion of the seriously considered studies.
- d. Reader is able to retrace the included and excluded studies anywhere in the article bibliography, reference, or supplemental source.
- If it satisfies 4 of the criteria = 4
If it satisfies 3 of the criteria = 3
If it satisfies 2 of the criteria = 2
If it satisfies 1 or 0 of the criteria = 1
- 6. Were the characteristics of the included studies provided?**
- a. In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions AND outcomes.
- b. Provide the ranges of relevant characteristics in the studies analyzed (e.g. race, sex, relevant, SES).
- c. The information provided appears to be complete and accurate (i.e. tolerable level of
- If it satisfies 3 of the criteria = 4
If it satisfies 2 of the criteria = 3
If it satisfies 1 of the criteria = 2
If it satisfies 0 of the criteria = 1

subjectivity).

- | | |
|---|--|
| 7. Was the scientific quality of the included studies assessed and documented? | If it satisfies 4 of the criteria = 4 |
| a. 'A priori' methods of assessment should be provided (experimental, quasi-experimental, natural designs). | If it satisfies 3 of the criteria = 3 |
| b. The scientific quality of the included studies appears to be meaningful. | If it satisfies 2 of the criteria = 2 |
| c. Discussion/recognition/awareness of level of evidence. | If it satisfies 1 or 0 of the criteria = 1 |
| d. Quality of evidence should be rated/ranked based on characterized instruments. (Characterized instrument is a created instrument that ranks the level of evidence, and Evaluation. | |
| | |
| 8. Was the scientific quality of the included studies used appropriately in formulating conclusions? | If it satisfies 4 of the criteria = 4 |
| a. The results of the methodological rigor and scientific quality should be considered in the analysis and conclusions of the review. | If it satisfies 3 of the criteria = 3 |
| b. The results of the methodological rigor and scientific quality are explicitly stated in formulating recommendations. | If it satisfies 2 of the criteria = 2 |
| c. Conclusions converge on a applied consensus statement. | If it satisfies 1 or 0 of the criteria = 1 |
| d. This practical consensus supports or refutes current practice. | |
| | |
| 9. Were the methods used to combine the findings of studies appropriate? | If it satisfies 4 of the criteria = 4 |
| a. Statement of the criteria that were used to decide that the studies analyzed were similar enough to be pooled. | If it satisfies 3 of the criteria = 3 |
| b. For the pooled results, a test of heterogeneity should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I ²) | If it satisfies 2 of the criteria = 2 |
| c. Is there a recognition of the heterogeneity or the lack thereof | If it satisfies 1 or 0 of the criteria = 1 |
| d. If heterogeneity exist a "random effects model" should be used and/or the rationale of combining should be taken into consideration or stated explicitly. | |
| e. If homogeneity exist, author state a rationale or a statistical test. | |
| | |
| 10. Was the likelihood of publication bias (aka "file | If it satisfies 3 of |

| | |
|---|---------------------------------------|
| drawer” effect) assessed? | the criteria = 4 |
| a. Recognition of publication bias or file-drawer effect | If it satisfies 2 of the criteria = 3 |
| b. An assessment of publication bias should be include graphical aids (e.g. funnel plot, other available tests) | If it satisfies 1 of the criteria = 2 |
| c. Statistical test (e.g. Egger regression test) | If it satisfies 0 of the criteria = 1 |
| 11. Was conflict of interest stated? | If it satisfies 3 of the criteria = 4 |
| a. Statement of sources of support | If it satisfies 2 of the criteria = 3 |
| b. No conflict of interest. This is subjective and may require some deduction or searching. | If it satisfies 1 of the criteria = 2 |
| c. An awareness statement of support or conflict of interest in primary inclusion studies | If it satisfies 0 of the criteria = 1 |

RESULTS AND DISCUSSION

The final pool of studies comprised 18 meta-analyses inclusive of studies conducted between 1986 and 2014, representing 1,193 independent effect sizes. The median year of publication was 2007 and the range for year of publication was 28 years. The sample consisted of 10 articles and 8 unpublished dissertations. A complete list of study characteristics is presented in table 1. All studies except one included an overall mean effect size or sufficient data to calculate the overall effect size. Only one study reported an overall negative effect size. The overall effect sizes ranged from -.11 to 1.02 in magnitude.

R-AMSTAR scores ranged from 18 to 35 or from a grade of C to A. The mean score on the R-AMSTAR was 25.5(5.50). This indicates that the meta-analyses in mathematics education studies investigating technology integration received an average grade of B. Individual scale means were highest for the following three items: “Was an ‘a priori’ design provided?”, “Was a comprehensive literature search performed?”, and “Were the methods used to combine the findings of studies appropriate?”. These results suggest that the majority of studies examined adhered to the most common methodological quality guidelines. While individual scale means were lowest for the following items: “Was the status of publication (i.e. grey literature) used as an inclusion criterion?”, “Was the scientific quality of the included studies used appropriately in formulating conclusions?”, and “Was conflict of interest stated?”. These data indicate that more work is needed to better utilize grey literature in the inclusion and exclusion process. Additionally, researchers must begin to consider the scientific quality of the included studies in the interpretation of meta-analysis results. The complete categorical item statistics from the R-AMSTAR are provided in table 3.

Between groups mean scores were calculated for the categorical variables publication status and technology intervention type. Articles had a mean score of 26.2(6.5), while unpublished dissertations had a mean score of 22.4(3.2). Plots of 95% confidence intervals indicate that the means difference between published and unpublished studies was not

statistically significantly different. The plots presented in figure 2 suggest that the confidence bands overlap substantially, thus the two groups are not statistically significantly different. Additionally, the differences across technology intervention types were also assessed. The mean scores in order from largest to smallest were: combination 28.3(5.5), software 26.12(6.3), computer assisted instruction 21.8(2.1), and calculator 19.2(1.6). Figure 4 presents the 95% confidence interval plots for technology type. The overlap between confidence bands indicates an absence of statistically significant differences between intervention types. Finally, to assess changes in study quality over time the bivariate correlation between publication year and R-AMSTAR score was analyzed. The correlational analysis revealed a positive relationship between year of publication and R-AMSTAR score, $r = .59$, $p < .001$. This indicates that as publication year increased the quality of studies also increased.

Table 2. Study characteristics

| Citation | Purpose | Source | <i>k</i> | <i>ES</i> | R-AMSTAR/Grade |
|--------------------------|--|--------|----------|-----------|----------------|
| Hembree & Dessart (1986) | Integrate the findings of the research on effects on students of using calculators in learning mathematics in Grades K-12. | J | 29 | .64 | 18.0/C |
| Lee (1990) | Determine the Effectiveness of CAI in elementary and secondary instruction. | D | 243 | .38 | 20.0/C |
| Chen (1994) | Synthesize and extract the main findings from studies on computer-based instruction (CBI) in mathematics education. | D | 76 | .50 | 19.5/C |
| Chadwick (1997) | Examine the effects of CAI in the secondary mathematics classroom on cognitive and affective outcomes. | D | 41 | .51 | 24.5/B |
| King (1997) | Determine the effect of computer-enhanced instruction (CEI) on college level mathematics. | D | 30 | .20 | 22.0/C |
| Nickolau (2001) | Synthesize the effects of hand-held calculators on K-12 mathematics achievement. | D | 24 | .54 | 18.5/C |
| Hsu (2003) | Examined the effectiveness of Computer Assisted Instruction (CAI) instruction in statistics education | D | 25 | .43 | 22.0/C |

| | | | | | |
|--|---|---|-----|------|--------|
| Ellington (2006) | Examined the effects of calculator use on student achievement and attitude levels | J | 54 | NA | 21.0/C |
| Schenker (2007) | Examine the effectiveness of using technology to enhance statistics instruction. | D | 117 | .24 | 28.0/B |
| Tokpah (2008) | Examined the Effects of Computer Algebra systems (CAS) on mathematics achievement. | D | 102 | .38 | 25.0/B |
| Rosen & Salomon (2007) | Examined the effectiveness of constructivist technology intensive learning environments versus traditional learning environments. | J | 32 | .46 | 25.0/B |
| Wang, Jiao, Young, Brooks & Olson (2007) | Examined the effect of testing mode (computer vs. paper and pencil) on mathematics achievement. | J | 14 | -.11 | 28.0/B |
| Li & Ma (2010) | Examined the effects on computer Technology on mathematics achievement in K-12. | J | 46 | .28 | 35.0/A |
| Larwin & Larwin (2011) | Determine the effectiveness of CAI student mathematics achievement in post-secondary statistics courses. | J | 219 | .57 | 23.0/B |
| Cheung & Slavin (2013) | Examined the effects of educational technology on mathematics achievement in K-12 settings. | J | 74 | .16 | 35.0/A |
| Steenbergen-Hu & Cooper (2013) | Examined the effects of intelligent tutoring systems on K-12 mathematics achievement. | J | 26 | .09 | 34.0/A |
| Moyer-Packenham & Westenskow (2013) | Synthesize the findings examining the effects of virtual manipulatives on student achievement. | J | 32 | .35 | 19.5/C |
| Chan & Leung (2014) | Evaluate the effects of digital geometry software on mathematics achievement. | J | 9 | 1.02 | 23.0/B |

Table 3. R-AMSTAR categorical item mean scores

| R-AMSTAR ITEM Category | <i>M</i> | <i>SD</i> |
|--|-----------------|------------------|
| 1. Was an 'a priori' design provided? | 3.72 | .46 |
| 2. Was there duplicate study and data extraction? | 2.28 | 1.49 |
| 3. Was a comprehensive literature search performed? | 3.28 | .75 |
| 4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? | 1.17 | 1.1 |
| 5. Was a list of studies (included and excluded) provided? | 1.56 | .62 |
| 6. Were the characteristics of the included studies provided? | 2.72 | .90 |
| 7. Was the scientific quality of the included studies assessed and documented? | 1.72 | 1.13 |
| 8. Was the scientific quality of the included studies used appropriately in formulating conclusions? | 1.17 | .71 |
| 9. Were the methods used to combine the findings of studies appropriate? | 3.22 | 1.26 |
| 10. Was the likelihood of publication bias (aka "file drawer" effect) assessed? | 1.94 | 1.2 |
| 11. Was conflict of interest stated? | 1.39 | .98 |

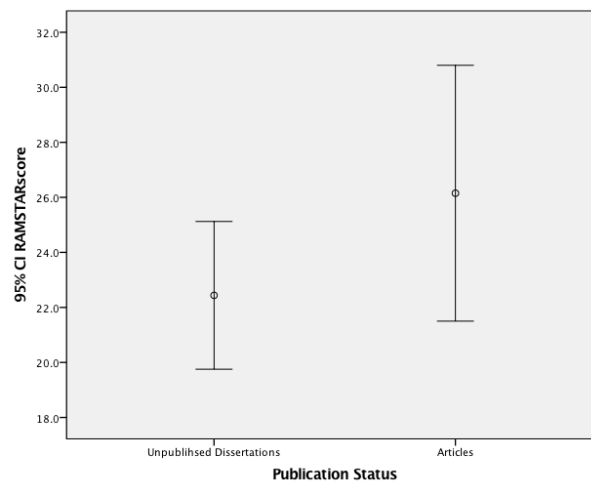


Figure 2. 95% Confidence Intervals for Publication Status

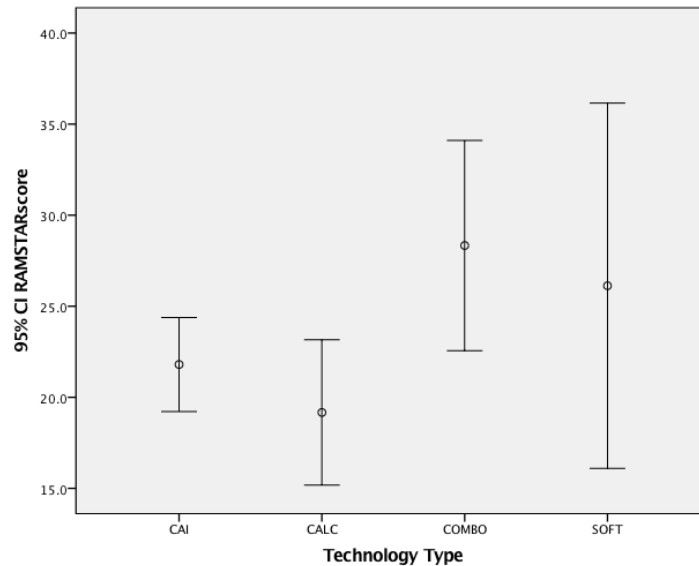


Figure 3. 95% confidence intervals for technology type

CONCLUSION

This study provides a comprehensive systematic review and critique of the design and reporting of meta-analytic research concerning technology integration in the mathematics classroom. The results of this study have several implications for research and instructional practice.

Mathematics educational researchers and other social scientist should consider ways to improve reporting practices and design quality. The available meta-analyses concerning technology integration in the mathematics classroom are relatively reliable and rigorous. However, to improve study quality and to promote meta-analytic thinking, researchers should consider adopting the Meta-Analysis Reporting Standards (MARS) at a minimum to ensure that the basic design and reporting considerations are addressed.

Mathematics educators should consider using a variety of technological tools in the mathematics classroom. Prior meta-analyses indicate that the effects of technology in the mathematics classroom vary across intervention, but are overall positive. Different technologies have specific instructional affordances and constraints. Given this challenge it is important that teachers consider a variety of tools to address different content and instructional classroom needs.

Based on the summary of almost 30 years of research this study provides important conclusions related to the reporting and methodological quality of studies in this area. The results of this critique suggest that the majority of meta-analysis reviewed score relatively well, and on average earned a grade of B. This suggest that most studies represent an above average level of design rigor and transparent reporting. However, several major reporting concerns were also observed across the majority of the studies. These concerns will effect the validity and reliability of reported summary effect sizes; thus they must be considered as the field continues to move forward. Reporting quality did not vary by publication status or technology intervention. This is important because it indicates that the qualities of many meta-analytic dissertations are comparable to published research. Additionally, because

design quality was not influenced by technology intervention the results across interventions are relatively similar regarding the design and reporting quality. Finally, the study substantiates the claim that as methods and reporting standards are increasing as researchers are conducting more sound meta-analyses in mathematics education research in this area. This is substantiated by the positive relationship between R-AMSTAR score and year of publication. In conclusion, as the field of mathematics education moves forward, technology will evolve as well, thus we as researchers and consumers of research must have valid and reliable data to make instructional and classroom decisions. It is my hope that this study informs current and future research towards this end.

REFERENCES

- American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33–40.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.
- Capraro, R.M. (2004). Statistical significance, effect size reporting, and confidence intervals: Best reporting strategies. *Journal for Research in Mathematics Education*, 35(1), 57–62.
- Chadwick, D.K.H. (1997). *Computer-assisted instruction in secondary mathematics classrooms: A meta-analysis* (Ed.D.). Drake University, United States-Iowa. Retrieved from <http://search.proquest.com/pqdtglobal/docview/304389928/abstract/9D8837D0D8EB4A14PQ/11>
- Chan, K. K., & Leung, S.W. (2014). Dynamic geometry software improves mathematical achievement: Systematic review and meta-Analysis. *Journal of Educational Computing Research*, 51(3), 311–325.
- Chen, T.Y. (1994). *A meta-analysis of effectiveness of computer-based instruction in mathematics* (Ph.D.). The University of Oklahoma, United States - Oklahoma. Retrieved from <http://search.proquest.com/pqdtglobal/docview/304105627/abstract/9D8837D0D8EB4A14PQ/5>
- Cheung, A.C.K., & Slavin, R.E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.
- Cooper, H. (2016). *Research synthesis and meta-analysis: A step-by-step approach* (Vol. 2). Sage publications.
- Cooper, H., & Patall, E.A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*, 14(2), 165-176.
- DeCoster, J. (2004). *Meta-analysis notes*. Retrieved from <http://www.stathelp.com/Meta%20analysis%202009-06-01.pdf>

- Ellington, A.J. (2006). The effects of non-CAS graphing calculators on student achievement and attitude levels in mathematics: A Meta-Analysis. *School Science and Mathematics*, 106(1), 16-23.
- Hembree, R., & Dessart, D.J. (1986). Effects of hand-held calculators in precollege mathematics education: a meta-analysis. *Journal for Research in Mathematics Education*, 17, 83-99.
- Hsu, Y. (2003). *The effectiveness of computer-assisted instruction in statistics education: A meta-analysis* (Ph.D.). The University of Arizona, United States - Arizona. Retrieved from <http://search.proquest.com/pqdtglobal/docview/305338759/abstract/9D8837D0D8EB4A14PQ/29>
- King, H.J. (1997). *Effects of computer-enhanced instruction in college-level mathematics as determined by a meta-analysis* (Ph.D.). The University of Tennessee, United States - Tennessee. Retrieved from <http://search.proquest.com/pqdtglobal/docview/304373727/abstract/9D8837D0D8EB4A14PQ/2>
- Larwin, K., & Larwin, D. (2011). A meta-Analysis examining the impact of computer-assisted instruction on postsecondary statistics education: 40 years of research. *Journal of Research on Technology in Education*, 43(3), 253-278.
- Lee, W.C. (1990). *The effectiveness of computer-assisted instruction and computer programming in elementary and secondary mathematics: A meta-analysis* (Ed.D.). University of Massachusetts Amherst, United States - Massachusetts. Retrieved from <http://search.proquest.com/pqdtglobal/docview/303852079/abstract/9D8837D0D8EB4A14PQ/18>
- Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review*, 22(3), 215-243.
- Moyer-Packenham, P. S., & Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning: *International Journal of Virtual and Personal Learning Environments*, 4(3), 35-50.
- National Council of Teachers of Mathematics. (2000). *Principles and standards of school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2008). Position statement: The role of technology in the teaching and learning of mathematics. Retrieved from <http://www.nctm.org/Standards-and-Positions/Position-Statements/Technology-in-Teaching-and-Learning-Mathematics/>
- Nikolaou, C. (2001). *Hand-held calculator use and achievement in mathematics: A meta analysis* (Ph.D.). Georgia State University, United States - Georgia. Retrieved from <http://search.proquest.com/pqdtglobal/docview/304696658/abstract/9D8837D0D8EB4A14PQ3>
- Rosen, Y., & Salomon, G. (2007). The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: A meta-analysis. *Journal of Educational Computing Research*, 36(1), 1-14.

- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev ed.). Beverly Hills, CA: Sage.
- Sanchez-Meca, J., & Marin-Martinez, F. (1998). Testing continuous moderators in meta-analysis: A comparison of procedures. *British Journal of Mathematical & Statistical Psychology*, 51, 311-326.
- Schenker, J.D. (2007). *The effectiveness of technology use in statistics instruction in higher education: A meta-analysis using hierarchical linear modeling* (Ph.D.). Kent State University, United States -Ohio. Retrieved from <http://search.proquest.com/pqdtglobal/docview/304835263/abstract/9D8837D0D8EB4A14PQ/25>
- Schmidt, F.L., & Hunter, J.E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage publications.
- Schmucker, C., Bluemle, A., Briel, M., Portalupi, S., Lang, B., Motschall, E., ... & Meerpohl, J. J. (2013). A protocol for a systematic review on the impact of unpublished studies and studies published in the gray literature in meta-analyses. *Systematic Reviews*, 2, 24-24.
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K-12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970-987.
- Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, 31, 25-32.
- Tokpah, C.L. (2008). *The effects of computer algebra systems on students' achievement in mathematics* (Ph.D.). Kent State University, United States - Ohio. Retrieved from <http://search.proquest.com/pqdtglobal/docview/304549974/abstract/9D8837D0D8EB4A14PQ>
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007). A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, 67(2), 219-238.
- Young, J. R., Ortiz, N., & Young, J. L. (2017). STEMulating interest: A meta-analysis of the effects of out-of-school time on student STEM interest. *International Journal of Education in Mathematics, Science and Technology*, 5(1), 62-74.
- Young, J.R., Young, J.L., & Hamilton, C. (2013). The Use of Confidence Intervals as a Meta-Analytic Lens to Summarize the Effects of Teacher Education Technology Courses on Preservice Teacher TPACK. *Journal of Research on Technology in Education*, 46(2), 149-172.
- Young, J.R. (2016). Unpacking TPACK in Mathematics education research: A systematic review of meta-analyses. *International Journal of Educational Methodology*, 2(1), 19-29. doi: 10.12973/ijem.2.1.19
- Young, J.R., & Young, J.L. (2012). "But that's not fair": Teacher technology readiness and African American Students'. *The Journal of the Texas Alliance of Black School Educators*, 4(1), 19-32.
- Zientek, L.R., Capraro, M.M., & Capraro, R.M. (2008). Reporting practices in quantitative teacher education research: One look at the evidence cited in the AERA Panel Report. *Educational Researcher*, 37, 208-216.