

## A study of meta-analyses reporting quality in the large and expanding literature of educational technology

Rana M. Tamim  
Zayed University

Evgueni Borokhovski, Robert M. Bernard, Richard F. Schmid, Philip C. Abrami, David I. Pickup  
Concordia University

As the empirical literature in educational technology continues to grow, meta-analyses are increasingly being used to synthesise research to inform practice. However, not all meta-analyses are equal. To examine their evolution over the past 30 years, this study systematically analysed the quality of 52 meta-analyses (1988–2017) on educational technology. Methodological and reporting quality is defined here as the completeness of the descriptive and methodological reporting features of meta-analyses. The study employed the Meta-Analysis Methodological Reporting Quality Guide (MMRQG), an instrument designed to assess 22 areas of reporting quality in meta-analyses. Overall, MMRQG scores were negatively related to average effect size (i.e., the higher the quality, the lower the effect size). Owing to the presence of poor-quality syntheses, the contribution of educational technologies to learning has been overestimated, potentially misleading researchers and practitioners. Nine MMRQG items discriminated between higher and lower average effect sizes. A publication date analysis revealed that older reviews (1988–2009) scored significantly lower on the MMRQG than more recent reviews (2010–2017). Although the increase in quality bodes well for the educational technology literature, many recent meta-analyses still show only moderate levels of quality. Identifying and using only best evidence-based research is thus imperative to avoid bias.

### *Implications for practice or policy:*

- Educational technology practitioners should make use of meta-analytical findings that systematically synthesise primary research.
- Academics, policymakers and practitioners should consider the methodological quality of meta-analyses as they vary in reliability.
- Academics, policymakers and practitioners could avoid misleading bias in research evidence by using the MMRQG to evaluate the quality of meta-analyses.
- Meta-analyses with lower MMRQG scores should be considered with caution as they seem to overestimate the effect of educational technology on learning.

*Keywords:* meta-analysis, systematic review, reporting quality, bias, educational technology, research methodology

## Introduction

Interest in applications of technology to education has been around for at least 70 years, and researchers have tried to document their effectiveness for all this time. In the 1950s–1960s, the emergence of television as a new medium of instruction, deemed ideal for distance education initiatives of the time, created a flurry of research that compared learning environments using interactive television with traditional technology-free classroom instruction (e.g., Carpenter & Greenhill, 1955, 1958). Similarly, various forms of computer-based instruction (1970s–1980s; e.g., Kulik & Kulik, 1987, 1991), multimedia (1980s–1990s; e.g., Maki & Maki, 2002), teleconferencing (1990s; e.g., McGreal, 1994) and, more recently, technology-based simulation, gaming and a host of Internet-based applications (e.g., D’Angelo et al., 2014) have been investigated from a comparative perspective in an attempt to judge their relative effectiveness. Inevitably, with all of this research activity, the need to summarise and evaluate the overall effectiveness of educational technology arose fairly early and persists to this day. With the number of primary research studies growing exponentially, especially in the field of rapidly evolving educational technology, practitioners’ reliance on research summaries increases proportionally.

While there have been some reviews of the narrative type (including of qualitative research), most reviews have been meta-analyses (Bethel & Bernard, 2010), conducted to synthesise the primary literature of the day so as to offer a quantitative estimate of the effectiveness of various technologies used for instruction and learning. From the late 1980s, there has been a fairly steady flow of such meta-analyses, and the end is not in sight as new manifestations of educational technology are developed and subjected to the scrutiny of primary researchers and reviewers. The majority of available meta-analyses have focused on learning achievement outcomes. These syntheses are necessary, as both researchers and practitioners recognise that a single or small set of studies, no matter how well conducted, cannot provide a robust comprehensive picture of whether/why or how an intervention is effective and useful. There is a significant challenge, however, in that not all meta-analyses are created equal. Indeed, some are quite poor, such that the uninformed consumer may assess their value far beyond what is justified. It is thus self-evident that both research progress and evidence-based practice should rely on high-quality syntheses, as the potential bias of poorly conducted and/or reported syntheses can actually mislead or impede.

To better understand and assess the quality of a meta-analysis, it is important to highlight what constitutes a good one. Meta-analysis dates back to the original work by Glass (1976), who introduced it as an alternative to narrative and vote-count reviews (Jackson, 1980). Since then, many general prescriptions for conducting and reporting meta-analyses have come into existence (e.g., Cooper, 2017; Lipsey & Wilson, 2001). There is also a growing literature on specific methodological issues, including research quality (e.g., Abrami & Bernard, 2012; Ahn et al., 2012), statistical analysis (e.g., Borenstein et al., 2009; Hedges & Olkin, 1985; Schmidt & Hunter, 2015), literature acquisition (e.g., Kugley et al., 2017; Pickup et al., 2018), quality assessment of primary studies (e.g., Cheung & Slavin, 2016; Valentine & Cooper, 2008), control for publication bias (e.g., Polanin et al., 2016), outlier identification (e.g., Viechtbauer & Cheung, 2010), dependency issues (e.g., Scammacca et al., 2014) and sources of confounding (e.g., Clark, 1985; Lipsey, 2003; Valentine & Thompson, 2013).

Quality of review methodology has been an issue since Slavin (1986) introduced the idea of a best evidence synthesis, one that uses only the highest quality evidence available around a particular question. Sometimes this means reviews that include only longitudinal experimental designs, studies with standardised measures, or studies that look beyond a single classroom. However, in other circumstances it means whatever the best evidence purview is of a particular question. Although Slavin's views have not become the standard, organisations such as the Campbell Collaboration (<https://campbellcollaboration.org/>) and the What Works Clearinghouse (Institute of Education Sciences, n.d.) have strived to raise the bar of the quality of evidence included in reviews in social sciences, in much the same the same way the Cochrane Collaboration (<https://www.cochrane.org/>) has done in the health sciences.

Like any other complex form of research, meta-analyses can be biased in a number of ways and, as a result, possibly misrepresent the state of research. There are often disconnects between the methodological literature (i.e., what should be done), methodological practices that populate the research literature (i.e., what is actually being done) and the public documentation of the research (i.e., what is reported in the literature). It comes down to this: When several, similar meta-analyses disagree, which results should be trusted most and which should be interpreted with caution or even rejected outright because of severe reporting bias? The use of the phrase "reporting bias" comes from the reality that a meta-analysis can only be evaluated based on the information provided by its authors. After all, if a journal publishes a meta-analysis with inadequate or confusing information, the readers cannot know with certainty if there are serious flaws in the review itself or in the reported outcomes.

Considering that meta-analyses are always retrospective – relying on research already conducted – the meta-analyst can do very little to improve the quality of primary studies under review, but instead is obligated to apply substantial efforts to examine them. For example, when inaccuracies in primary research are detected, meta-analysts may decide to exclude methodologically flawed research from consideration. Otherwise, they may decide to explore and document the sources of error, while following the rules of systematicity and transparency in reporting both rationale and outcomes for every decision made. Failure to do either would result in a biased meta-analysis. This being said, it is important to note that bias is not a unilateral aspect of a given meta-analysis but an outcome of one or more errors resulting from inadequate processes employed in data collection, manipulation, analysis, interpretation and/or presentation of research findings (Bernard, 2014). We consider flaws in any of these processes to be a methodological reporting bias that could potentially lead to presenting an inaccurate picture of the past research.

As such, the main objective of this study was to investigate the methodological reporting quality of a large sample of meta-analyses addressing educational technology while answering the following questions:

- (1) Can the amount of potential reporting bias in a given meta-analysis be reasonably assessed to inform readers of the credibility of its results?
- (2) Do meta-analyses with higher methodological reporting quality yield outcomes that differ from poorer quality ones. If so, in what ways?
- (3) What are the primary sources of reporting bias, and do they influence the overall findings of these meta-analyses?
- (4) Has reporting quality changed over the years 1988–2017?
- (5) Are advances in meta-analysis methods over the years reflected in changes in the reporting quality of educational technology meta-analyses?

## **Method**

For the purpose of the project, we used the Meta-analysis Methodological Reporting Quality Guide (MMRQG), a 22-item instrument that we developed and designed to assess reporting bias. We tested the first version of the MMRQG using studies drawn from a second-order meta-analysis addressing technology integration (Tamim et al., 2011). We further refined the instrument in a subsequent follow-up with studies from Bernard et al. (2009) and Schmid et al. (2014). The final form of the MMRQG is presented in Table 1. The items are coupled with brief descriptions in the form of guiding questions that allow users to score each aspect of a meta-analysis indicating whether it conforms to rigorous standards of meta-analytical research. For every item, the information presented in a manuscript can be used to provide a score based on a three-level qualitative characterisation. Each score is decided according to the response to the corresponding guiding question. The three levels are as follows:

- The score 0 is given when the report does not provide any relevant information with regards to the guiding question. This in turn indicates that the requirements for proper meta-analytical procedures and/or decisions are not met.
- The score 1 is given when the report provides limited information that does not fully address the guiding question. This in turn indicates that the requirements for proper meta-analytical procedures and/or decisions are partly met.
- The score 2 is given when the report provides explicit and comprehensive information that addresses the guiding question. This in turn indicates that the requirements for proper meta-analytical procedures and/or decisions are fully met.

Table 1  
*MMRQG: Items and descriptions*

Items	Guiding questions
1 Research question	Is the research objective and/or the question clearly stated?
2 Contextual positioning of the research problem	Is the rationale for meta-analysis adequate, conceptually relevant, and supported by empirical evidence (i.e., the quality and relevance of the literature review section)?
3 Time frame	Is the time frame defined and adequately justified in the context of the research question and prior reviews?
4 Experimental group	Is the experimental group clearly defined and described in detail (possibly with examples)?
5 Control group	Is the control group clearly defined and described in detail (possibly with examples)?
6 Outcomes	Are the measures of the identified outcome(s) – dependent variables – appropriate and relevant to the research question and sufficiently described?
7 Inclusion criteria	Are the inclusion criteria clearly stated and described in detail (possibly supported by examples from the reviewed literature)?
8 Targeted literature	Is the targeted literature exhaustive and includes all types of published and unpublished literature?
9 Resources used	Are the resources used to identify relevant literature representative of the field and exhaustive (i.e., do they include multiple electronic databases, hand searches, branching)?
10 Search strategy	Is the list of search terms provided and appropriate for each individual source (e.g., modifying key words for specific databases)?
11 Article review	Is the article review process implemented by two or more researchers, working independently, with reasonable inter-rater reliability levels?
12 Effect size extraction	Do two or more researchers with reasonable inter-rater reliability levels implement the independent effect size extraction process?
13 Study feature coding	Do two or more researchers implement the study feature independent coding process with reasonable inter-rater reliability?
14 Validity of included studies	Are all aspects of validity of included primary studies explicitly discussed, defined and consistently addressed across studies?
15 Independence of data	Is the issue of dependency among included studies addressed with methods for assuring data (i.e., samples and outcomes) independence are appropriate and adequately described?
16 Effect size metrics and extraction procedures	Are the effect size metrics and extraction procedures used appropriate and fully described including necessary transformations?
17 Publication bias	Are procedures for addressing publication bias adequately substantiated and reported?
18 Treatment of outliers	Are criteria and procedures for identifying and treating outliers adequately substantiated and reported?
19 Overall analyses	Is the overall analysis performed according to standard procedures (e.g., correct model use, homogeneity assessed, standard errors reported, confidence intervals reported)?
20 Moderator variable analyses:	Are moderator variable analyses performed according to the proper analytical model and is appropriate information reported (e.g., <i>Q</i> -between, test statistics provided)?
21 Reporting results	Are the appropriate statistics supplied for all analyses and explained in enough detail that the reader will understand the findings?
22 Appropriate interpretation	Are the findings summarised and interpreted appropriately in relation to the research question?

### Search for and review of meta-analyses

To address the research questions about the relationship between meta-analysis quality and outcomes, we assembled 52 studies that matched the central inclusion criteria of meta-analysis and technology integration in education inclusive of primary, secondary, post-secondary, higher education and vocational education. Our first compilation was derived from the original meta-analyses obtained from a second-order meta-analyses (Tamim et al., 2011). This yielded 38 reviews of technology integration in education dating from 1988 to the end of 2008. Next, we conducted search updates to locate additional meta-analyses published from 2008 to 2017. We used the same search strategies and inclusion and exclusion criteria from Tamim et al. (2011) to provide consistency across the collection. We accessed the following data sources:

- electronic searches using major databases: ERIC, PsycINFO, Education Index, PubMed (Medline), AACE Digital Library, British Education Index, Australian Education Index, ProQuest Dissertations and Theses Full-text, EdITLib, Education Abstracts and EBSCO Academic Search Complete
- web searches using Google and Google Scholar
- manual searches of major journals, including *Review of Educational Research*, *Computers in Education* and an array of other journals devoted to publishing reviews of research
- reference lists of prominent articles and major literature reviews.

The search strategy used the term *meta-analysis* and its synonyms (e.g., *quantitative review*, *systematic review*). In addition, we used search terms relating to computer technology use within educational contexts. These varied according to the specific descriptors within different databases. Generally, they included terms such as *computer-based instruction*, *computer-assisted instruction*, *computer-based teaching*, *electronic mail*, *information communication technology*, *technology uses in education*, *electronic learning*, *hybrid courses*, *blended learning*, *teleconferencing*, *Web-based instruction*, *technology integration* and *integrated learning systems*. The complete list of included meta-analyses are available upon request, while reviews cited in this manuscript are included in the reference list and marked by an asterisk.

The inclusion criteria required that a meta-analysis address the effectiveness on learning (i.e., achievement outcomes) from instruction that used educational technology in the experimental condition in comparison with technology-free instruction in the control condition. Meta-analyses of distance education and focused on a single tool or application (e.g., single branded software packages) were excluded. Even though there are reviews from the same first author(s), care was taken to ensure that a different issue or technology was addressed in each (e.g., Cheung and Slavin are represented twice, once for a meta-analysis on K-12 reading (2012) and another (2013) on K-12 mathematics). In cases of multiple reviews conducted by the same authors on the same topic and building upon previous collections, only the final cumulative meta-analysis was admitted (e.g., Kulik & Kulik, 1991).

As a result of the update, 14 meta-analyses were added to the original collection of 38, bringing the total to 52. There were two meta-analyses that, in addition to technology-free control conditions, explored other types of comparisons (e.g., more vs less technology or the added value of specific pedagogical approaches to technology-based instruction), but since the outcomes of these two types were reported separately for different types of comparisons, these meta-analyses were partly retained. Specifically, in Schmid et al. (2014), one part was excluded from consideration because it dealt exclusively with comparisons between technology-rich and technology-lean conditions. Likewise, in D'Angelo et al. (2014) only simulation versus no simulation comparisons were included, separated from the collection that synthesised simulations versus simulations with various enhancements.

### Data extraction and coding of meta-analyses included

Two expert reviewers (i.e., two of us) worked independently to assess each included meta-analysis using the 0 to 2 scale described earlier. They then compared their coding results. Disagreements were marked, discussed and resolved. Inter-rater agreement rate for initial independent coding was 93.5% (Cohen's  $\kappa = 0.87$ ).

Simple arithmetical sums of all codes across 22 MMRQG items formed the total index of methodological quality with a theoretical range from 0 (all items coded "no") to 44 (all items coded "yes"). The resulting scale appears to be a reasonable approximation of a continuous scale, and as such, was employed in the subsequent analyses. Additionally, each individual item has three levels broadly reflecting low, moderate and high methodological quality of reporting.

A single weighted average achievement effect size was extracted from each meta-analysis. We use Hedges'  $\bar{g}$ , with the bar above  $g$ , to indicate the average effect size across all the studies included in a given meta-analysis. In all cases, this was the overall test of the effect of the technology being reviewed (i.e., the effect of technology versus no technology on learning achievement). Effect sizes were extracted as reported and then converted (when required) to the common metric Hedges'  $\bar{g}$  (i.e., a variant of Cohen's  $\bar{d}$  corrected for small sample bias). The standard error of each meta-analysis was estimated based

on the number of effect sizes reported rather than the sample size (i.e., number of participants), as is customary in individual meta-analyses.

In synthesising studies and in moderator variable analysis, the average effect size of each meta-analysis was weighted ( $W_{Random}$ ) by the inverse of its variance, which in the random effects model includes the sum of both within-study ( $V_{\bar{g}} = SE_{\bar{g}}^2$ ) and average between-study variance tau-squared ( $\tau_{\bar{g}}^2$ ).

## Statistical analysis and results

In total, 52 meta-analyses were located, reviewed and analysed. The metric resulting from all syntheses of the average of  $k$  studies is symbolised as  $\bar{g}^+$ , the average of the averages. Weighted average effect size was used as the outcome measure in these analyses because this metric is generally considered the bottom line when educational practitioners and policymakers examine the findings of a meta-analysis (see Cheung & Slavin, 2016). This average is considered a standardised index of the effectiveness of instructional interventions compared with an alternative control condition.

The main purpose of this study was to determine whether the methodological quality of reporting of meta-analyses in the educational technology literature predicts average study findings. The grand mean of the MMRQG items for all reviews was 0.927, close to the theoretical average of 1.0, with a standard deviation of 0.424.

One of the first questions we asked before proceeding to a more in-depth analysis was: Can this collection of meta-analyses, all addressing roughly the same question, be synthesised? Two analyses, both involving publication source, were conducted to answer this question. First, methodological quality equivalence was tested across the publication sources: journal articles, dissertations, reports and conference proceedings. Then, publication sources were tested for equivalence of their respective effect sizes. There were 37 journal articles, 8 dissertations, 4 reports and 3 conference proceedings.

An ANOVA conducted on MMRQG ratings across publication sources produced an overall significant effect ( $F[1, 3] = 3.075, p = .04$ ), but individual paired differences were not found in subsequent post hoc analyses. However, the means followed the pattern that more elaborate reporting in dissertations ( $M = 1.25$ ) and research reports ( $M = 1.45$ ), where there are fewer restrictions on length and the depth of analysis, might differ from other reporting sources where length is more of an issue – journal articles ( $M = 0.86$ ) and conference proceedings ( $M = 0.64$ ), despite their peer-reviewed nature, which might be seen as a reason for serious concern. With respect to the distribution of average effect sizes across publication sources, we found no significant differences:  $Q$ -between, ( $df = 2$ ), = 2.10,  $p = .55$ . As a result, we considered this collection to be reasonably homogeneous.

## Quantitative synthesis of 52 meta-analyses

One of the first steps in this analysis was to examine the meta-analytic results of the entire collection of reviews in relation to their quality ratings. Since both effect sizes and the MMRQG results can be characterised as continuous scales, simple meta-regression was conducted to determine, initially, if there was a predictive association between the two sets of values. The inverse variance weighted method of moments (random effects) analysis yielded a significant regression result (coefficient  $\beta = -0.129, df = 1, z = -2.29, p = .022$ ). The negative sign indicates an inverse relationship between effect sizes and MMRQG average values (i.e., larger MMRQG averages or better-quality meta-analyses predict smaller average effect sizes and vice versa).

The second step was to examine this finding in greater detail, first by synthesising all 52 meta-analyses as a group to establish a baseline for further categorisation of the collection based on MMRQG scale values (Table 2) and then by testing between-group differences across groups of studies classified as high, medium and low quality on the MMRQG (Table 3).

Table 2  
Overall weighted average effect size and heterogeneity statistics

Population estimates	<i>k</i>	$\bar{g}^+$	<i>SE</i>	CI: Lower 95th	CI: Upper 95th
Random effects model	52	0.41*	0.03	0.34	0.47
Heterogeneity analysis: $Q$ -Total = 57.96 ( $df = 51$ ), $p = .23$ , $I^2 = 12.01$ , $\tau^2 = 0.006$					

\* $p < .001$ .

The fixed and random models produced similar results because of the relatively small degree of between-study heterogeneity found in the collection. The test of  $Q$ -Total is not significant and  $I^2$  is small. Therefore, average between-study variability, characterised as  $\tau^2$ , is also small. The study-weighted random effects average effect size of  $\bar{g}^+ = 0.41$  is significant and is considered an educationally relevant moderate average effect size (Cohen, 1988).

The next analysis took the form of a standard mixed effects moderator variable analysis, with the 52 meta-analyses divided into three quality categories based on their MMRQG scale averages. Studies were categorised by MMRQG average based on the mean and standard deviation of the entire collection of 52 studies ( $M = 0.93$ ,  $SD = 0.42$ ). Studies whose MMRQG averages exceeded +1.0 standard deviations above the average, or  $M \geq 1.35$  ( $K = 7$ ), were classified as high in methodological quality (the range was from 1.36 to 1.91, where 2.0 is the maximum score attainable). Studies whose averages ranged between 1.0 standard deviations below and above the mean ( $M < 0.50$  to  $< 1.34$ ) were considered to be of moderate methodological quality ( $K = 36$ ), and studies whose averages fell  $\leq 1.0$  standard deviations below the mean ( $K = 9$ ) were classified as lower in quality. Studies in this category ranged from averages of  $M = 0.27$  to  $M = 0.50$ . Table 3 shows the mixed effect moderator variable analysis associated with these classification categories. The categories were significantly different, based on a  $Q$ -between = 7.41,  $df = 2$ ,  $p = .03$ . In follow-up analysis, lower- and medium-quality studies were found to not differ ( $Q$ -between = 0.03,  $df = 1$ ,  $p = .86$ ), while higher-quality studies did differ from medium-quality studies ( $Q$ -between = 6.84,  $df = 1$ ,  $p = .009$ ) and were almost significantly different from lower-quality studies ( $Q$ -between = 3.30,  $df = 1$ ,  $p = .07$ ).

Table 3  
Mixed-effect moderator variable analysis for three categories of MMRQG quality

Quality categories	<i>k</i> (%)	$\bar{g}^+$	<i>SE</i>	CI: Lower 95th	CI: Upper 95th
Lower-quality studies ( $M < 0.50$ )	9 (17.3)	0.40*	0.05	0.18	0.55
Medium-quality studies ( $M > .50$ and $\bar{M} M < 1.35$ )	36 (69.2)	0.45*	0.37	0.39	0.54
Higher-quality studies ( $M \geq 1.35$ )	7 (13.5)	0.27*	0.05	0.18	0.37
Test of category differences: $Q$ -between = 7.51 ( $df = 2$ ), $p = .03$					

\* $p < .001$ .

This moderator variable analysis suggests that average effect sizes tend to be lower in higher methodological quality studies. This is in line with the findings from the above meta-regression. At a minimum, there seems to be some degree of reporting bias related to effect size in the lowest-quality category of meta-analyses and possibly the medium-quality reviews as well.

### The MMRQG results in detail

Table 4 displays the means and standard deviations of all 22 MMRQG items. Many of the item means were above or around the mean of the whole distribution ( $M = 0.927$ ,  $SD = 0.424$ ). Because of the significant overall relationships described in Table 3 between quality items and average effect size, all of the items on the MMRQG were tested individually as moderator variables (the categories were 0, 1 and 2) to determine which, if any, of the items discriminated between average effect sizes at the individual item level. The results of these analyses produced nine items and are shaded in Table 4 (these are the only items that significantly discriminated across effect sizes). In the case of all nine items, there is an inverse (or nearly inverse) relationship with the MMRQG scale, with codes 0 (low quality) and +1 (acceptable quality) generally producing higher effects than the code of +2 (high quality), suggesting that this group of items likely contributes to the findings of the meta-regression analysis.

Table 4  
*Means and standard deviations of 22 MMRQG items (k = 52)*

MMRQG items	<i>M</i>	<i>SD</i>
1. Research question	1.67	0.51
2. Contextual positioning of the research problem	1.37	0.66
3. Time frame	0.90	0.66
4. Experimental group	1.13	0.74
5. Control group	0.54	0.67
6. Outcomes specification	1.02	0.70
7. Inclusion criteria	1.52	0.50
8. Targeted literature	0.96	0.71
9. Resources used	1.23	0.67
10. Search strategy	0.79	0.70
11. Article review	0.21	0.61
12. Effect size extraction	0.29	0.64
13. Study feature coding	0.92	0.90
14. Validity of included studies	0.62	0.75
15. Independence of data	0.48	0.67
16. Effect size metrics and extraction procedures	1.25	0.71
17. Publication bias	0.58	0.87
18. Treatment of outliers	0.50	0.80
19. Overall analyses	1.02	0.61
20. Moderator variable analyses	0.92	0.68
21. Reporting results	1.17	0.68
22. Appropriate interpretation	1.31	0.64

Note. Grand *M* = 0.927, Grand *SD* = 0.424, High:  $\geq 1.35$ ; Low:  $\leq 0.50$ .

Shaded items discriminate across the 0, 1 & 2 scale points.

To test this further, these nine items were removed from the MMRQG leaving 13 items. Meta-regression was again run, and the results were  $\beta$  coefficient = -0.060,  $SE = 0.07$ ,  $df = 50$ ,  $t = -.086$ ,  $p = .40$ . The sign of the  $\beta$  coefficient remains negative, but the 13-item scale no longer correlates significantly with average effect size. By contrast, when these nine items were isolated and tested as a group in meta-regression, the result was a negative and significant slope (coefficient  $\beta = -0.154$ ,  $SE = 0.04$ ,  $df = 50$ ,  $p = .0008$ ). This is not proof of outcome bias or even reporting bias, of course, but it is highly suggestive that it is in these nine areas where improvement is needed in the educational technology meta-analysis literature and that insufficient or inadequate attention to these methodological aspects of meta-analysis implementation may distort findings and jeopardise their credibility.

Four of the items (Items 2, 4, 5 & 6) deal with defining or specifying elements in the early stages of conceptualising a meta-analysis, before any studies are located or data are acquired. Items related to publication reporting bias did not emerge as discriminating. However, Item 11, Reviewing articles, which has to do with the number of raters employed, the rigor of the review process and inter-rater reliability, did. Lower scores on the MMRQG for this item mostly relate to the meta-analyst using one rather than more than one independent coders in manuscript review and selection. Items 14 and 15 have to do with describing and applying consistent standards across primary studies admitted to a meta-analysis, in terms of their validity and the independence of extracted effect sizes. Item 18 involves examining the distribution of studies for outliers and other potential aberrations and making necessary adjustments, and Item 20 is related to descriptions of moderator variable analysis (e.g., using the appropriate analytical model).



Table 5  
Mixed-effects moderator variable analysis across MMRQG scale items

MMRQG scale values	<i>k</i>	$\bar{g}^+$	<i>SE</i>	CI: Lower 95th	CI: Upper 95th
<b>Item 2: Positioning of research problem contextually</b>					
0 (low reporting quality)	5	0.41*	0.10	0.21	0.61
1 (acceptable reporting quality)	23	0.52*	0.06	0.40	0.63
2 (high reporting quality)	24	0.31*	0.04	0.24	0.38
Test of category differences: <i>Q</i> -between = 9.36 ( <i>df</i> = 2), <i>p</i> = .001					
<b>Item 4: Defining the experimental group</b>					
0	11	0.56*	0.11	0.35	0.78
1	23	0.41*	0.05	0.31	0.51
2	18	0.30*	0.04	0.23	0.38
Test of category differences: <i>Q</i> -between = 6.36 ( <i>df</i> = 2), <i>p</i> = .04					
<b>Item 5: Defining the control group</b>					
0	29	0.45*	0.05	0.35	0.54
1	18	0.42*	0.06	0.31	0.52
2	5	0.25*	0.06	0.14	0.36
Test of category differences: <i>Q</i> -between = 7.27 ( <i>df</i> = 2), <i>p</i> = .03					
<b>Item 6: Specifying outcome measures</b>					
0	12	0.56*	0.08	0.40	0.71
1	27	0.31*	0.04	0.23	0.38
2	13	0.37*	0.07	0.24	0.50
Test of category differences: <i>Q</i> -between = 7.95 ( <i>df</i> = 2), <i>p</i> = .02					
<b>Item 11: Reviewing articles</b>					
0	46	0.44*	0.03	0.37	0.51
1	1				
2	5	0.28*	0.05	0.17	0.39
Test of category differences: <i>Q</i> -between = 5.84 ( <i>df</i> = 1), <i>p</i> = .02					
<b>Item 14: Establishing the validity of included articles</b>					
0	28	0.46*	0.05	0.37	0.55
1	16	0.41*	0.06	0.29	0.53
2	8	0.28*	0.05	0.19	0.38
Test of category differences: <i>Q</i> -between = 6.87 ( <i>df</i> = 2), <i>p</i> = .03					
<b>Item 15: Maintaining independence of effect sizes</b>					
0	32	0.49*	0.05	0.39	0.58
1	15	0.29*	0.06	0.17	0.40
2	5	0.30*	0.05	0.19	0.40
Test of category differences: <i>Q</i> -between = 9.74 ( <i>df</i> = 2), <i>p</i> = .01					
<b>Item 18: Dealing with outliers</b>					
0	36	0.44*	0.04	0.35	0.52
1	6	0.51*	0.08	0.36	0.66
2	10	0.28*	0.05	0.18	0.37
Test of category differences: <i>Q</i> -between = 8.90 ( <i>df</i> = 2), <i>p</i> = .01					
<b>Item 20: Performing moderator variable analysis</b>					
0	14	0.45*	0.07	0.31	0.59
1	28	0.46*	0.05	0.37	0.56
2	10	0.29*	0.04	0.20	0.38
Test of category differences: <i>Q</i> -between = 8.08 ( <i>df</i> = 2), <i>p</i> = .02					

\**p* < .001.

Note. *k* = 1 is reported for description only and not included in the *Q*-between analysis.

Since average effect size in a meta-analysis is the metric by which most consumers judge the strength of a treatment or intervention of interest, it is important that this metric be as accurate as possible. As is clear from Table 3, studies classified as lower and medium quality on the MMRQG can differ as much as 0.13 to 0.18 standard deviations, representing a 5% to 7% overestimate, from reviews with higher quality reporting standards with an average of 0.27 standard deviations.

### The moderating effect of publication date

Publication date is one of just a few common moderator variables that were extracted from these meta-analyses. It is important to this assessment, however, because of the relatively short history of the modern practice of synthesising studies using meta-analysis. The methodology literature began rather slowly at first after Glass (1976) introduced the idea and described basic principles, but by the mid-1980s, just before the first studies in this collection were published, works by methodologists such as Hedges and Olkin (1985) were beginning to expand and clarify statistical procedures and works by more general methodologists, such as Rosenthal (1984), Cooper (2017), Cooper et al. (1994) and Lipsey and Wilson (2000), were providing frameworks, details and methodological standards. The Campbell Collaboration came into being in this era (1999), providing meta-analysts with assistance and methodological guidance (Littell & White, 2018). More recent works by authors, such as Borenstein et al. (2009), revisited and expanded the statistical procedures previously addressed. All of this is to say that meta-analysis has seen a remarkable surge of methodological and procedural literature over its lifetime, as well as countless meta-analyses in research areas from education and the social sciences to the medical sciences. Therefore, a possible hypothesis is that time and the provision of an expanded knowledge base and tools account for some of the variance observed in this exercise.

To test the premise that methodological reporting has improved over time in this literature (as one would expect given the surge of methodological literature), simple linear regression was conducted to determine if there is a connection between the publication dates of these 52 meta-analyses and the MMRQG score calculated for each study. This result was expected to be positive. The results of this analysis are shown in Table 6 and a scatterplot is shown in Figure 1. As expected, the result was positive and substantial with an  $R$  of 0.512 and a corresponding  $R^2$  of 0.262. This means that more than one quarter of the variability associated with the MMRQG is accounted for by publication date, expressed as a continuous variable.

Table 6

*Results of regression analysis of publication Date and MMRQG scores*

Model	$\beta$	Standard error	Beta	$t$ value	Significance
Intercept	-63.88	15.29		-4.18	< .001
Publication date	0.03	0.01	0.51	4.23	< .001

*Note.* ANOVA: Regression = 17.91,  $p < .001$  (1, 51), Residual = 9.64, (50, 51), Total = 13.09.

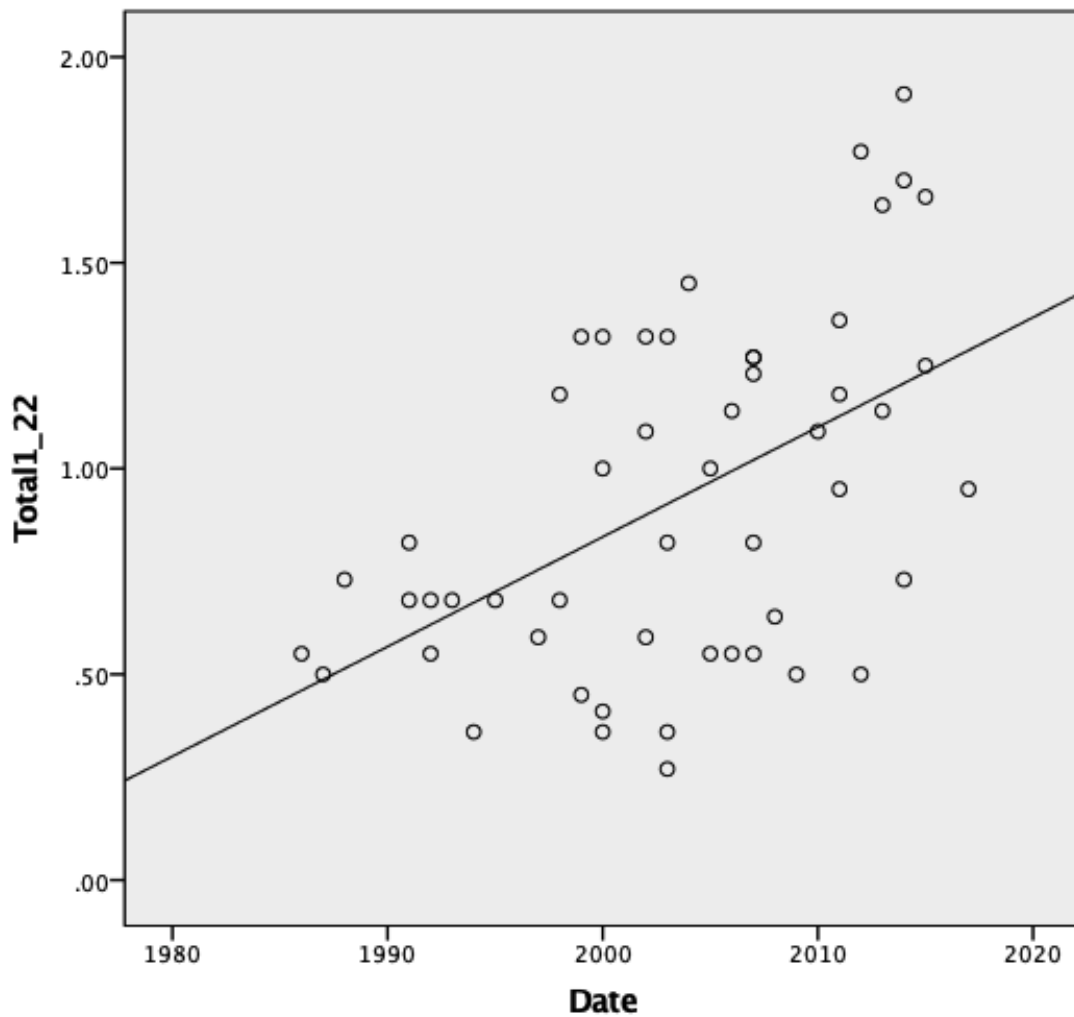


Figure 1. Scatter plot and regression line for publication date and total MMRQG scores

To further investigate if there is a particular period in which this shift occurred, publication date was classified as a categorical variable with three levels, 1988–1999, 2000–2009 and 2010–2017. The results yielded a significant difference among categories (see Table 7), and post hoc analysis (i.e., Tukey HSD) yielded a split between categories 1 and 2, grouped in one subgroup and category 3 in another. This result suggests that, on average, greater reporting rigor was applied in the most recent category (2010–2017), compared with the two previous ones (1988–1999 and 2000–2009). These results are consistent with what one would expect as a methodological literature that began in the 1970s, with little real guidance available, matured, became more accessible and more easily understood and applied. That said, variability in quality remained a prominent issue even in the most recent time period. Although between 1988 and 2009, only one study fell into the high-quality range, from 2010 to 2017, one meta-analysis was still deemed low quality, seven medium and six high. As such, researchers must continue to adhere to rigorous standards, and reviewers must insist upon the best that authors can provide when considering publication.

Table 7  
ANOVA of date categories and MMRQG scores

Source	SS	df	MS	F	Significance
Between groups	2.41	2	1.20	8.95	.001
Within groups	6.58	49	0.13		
Total	8.99	51			

Note. Post hoc comparisons revealed that Group 1 ( $N = 16$ ,  $M = 0.74$ ) is not significantly different from Group 2 ( $N = 23$ ,  $M = 0.85$ ), with both being significantly lower than Group 3 ( $N = 13$ ,  $M = 1.29$ ).

## Discussion

This study is an investigation of the reporting quality of meta-analyses produced in the educational technology literature between 1988 and 2017. We located and analysed 52 meta-analyses for methodological reporting quality using the MMRQG instrument we developed and pilot-tested. The premise underlying this study is that research evidence, as summarised in reviews, especially in meta-analyses, represents an important contribution to both educational practice and policymaking, and that the veracity and accuracy of the reporting in these meta-analyses is critical to best evidence-based progress within the field to aid the work of practitioners, policymakers and researchers alike. Equipping educators with the adequate means for assessing methodological reporting quality of meta-analyses in their respective areas of interest could be instrumental in ensuring the applied value of research evidence.

Our first step was to summarise the reviews as a whole. We found an overall average random effect size for 52 meta-analyses of  $\bar{g}^+ = 0.41$ ,  $p < .001$ , just slightly higher than that found by Tamim et al. (2011) for 25 meta-analyses. Though effects are quite consistent across the collection and fall under the category of low-to-moderate in Cohen's terms, the interpretation of these findings is not that straightforward. In simple regression analysis, we found an inverse relationship between MMRQG scores and effect size, meaning that lower quality scores predicted higher effect sizes and the reverse. Put simply, the data suggest that the contribution of technology to learning outcomes has been over-estimated. Relatively weaker meta-analyses consistently yield higher effect sizes, thus potentially misleading practitioners and policymakers. Whether any differences are systemically critical is a judgement call. In a field littered with hype and failed promises (Amiel & Reeves, 2008; Cuban & Jandric, 2015), stakeholders should rely upon only the best-quality evidence.

We then proceeded to examine individual MMRQG items to determine which were likely related to this inverse relationship. We found nine items in total where there was a significant inverse relationship with average subgroup effect size (i.e., treating the 0–2 scale as continuous). Five of these items were from the first half of the guide that deals, generally, with conceptualisation, the nature of the evidence and a description of the problem being addressed. They are stating and conceptualising the research problem (Item 2); defining the experimental (Item 4) and control groups (Item 5); specifying the outcome measures (Item 6); and reliability of reviewing articles (Item 11). The last four items were from the group of 12–22 of the guide which deals more with methodology: establishing study validity (Item 14); maintaining independence of effect sizes (Item 15); dealing with outliers (Item 18); and performing moderator variable analysis (Item 20). In all cases, the pattern was similar: lower quality or absent descriptions produced higher average effect sizes. We do not argue that these items themselves, either individually or collectively, produced higher or lower average effect sizes, but we maintain that they are indicators of lower or higher quality status and thus are significantly correlated with effect size (i.e., in the research literature the two often coincide).

In the last part of the Statistical analysis and results section, we examined MMRQG totals as they were distributed across the years 1988 to 2017 in three date categories. Here, the quality indicator sums were positively correlated with publication date. That is, the more recent the publication date, the higher the quality scale. In ANOVA, we found a significant overall  $F$  ratio among the three date categories. In post hoc analysis, Categories 1 (1988–1999) and 2 (2000–2009) were not significantly different from each other, but Category 3 (2010–2017) was significantly different from both of them. These findings bode well for the continued practice of synthesising quantitative, experimental studies in the educational technology literature, as the quality of such syntheses (and hence, their relevance and applied value) does improve. We assume that this result is partly explained by the wealth of methodological literature that has arisen, especially during the years of the last category. However, as noted above, more than half of the recent reviews still fell short of the high-quality designation. It is thus incumbent upon the journal review process to detect gaps or flaws in provided information. Failure to do so may result in poorer quality outcomes being propagated, thus leading to interpretative bias.

## Potential limitations

It must be recognised that there is no presumption of causation in these analyses. Therefore, the results are purely descriptive, as we had to rely on evidence supplied by the authors of the reviews in the journals, dissertations, research reports and conference proceedings. Although we assume that these sources

accurately portray the methodology used, it is of course possible that they do not. It is also possible that there is some amount of measurement error in the MMRQG, in spite of the fact that we are seasoned professionals with many years of experience conducting and writing about meta-analysis. The inter-rater reliabilities suggest that we largely agreed, after independently rating the reviews. These potential limitations, however, are not unlike the issues that all types of reviewers (i.e., journal reviewers, dissertation committees) face in every assessment of a meta-analysis.

### **The future of research and reviews of research in educational technology**

Four decades after its introduction, meta-analysis has become the most widely used methodology for summarising and synthesising quantitative experimental and correlational studies on which educational practitioners have to rely simply because of human impossibility to meaningfully process the overwhelming amount of data in primary empirical research. However, producing a meta-analysis is a time-consuming and expensive process that is highly susceptible to errors at all stages of its production unless undertaken with extreme care and rigor. Some are small and do not matter much to the final outcome. Some are so egregious that they can severely bias the finding and the message they are intended to deliver to the educational technology community. Although poor reporting standards are not necessarily indicative of errors in practice, they are the only evidence of study findings that are widely available to consumers. Of course, this is generally true for all forms of research in all fields, and because of this, replications (even of meta-analyses) are advisable. It is worth noting that a meta-analysis is in fact a form of compilation and summation of replications. Single studies may be flawed, but when taken together, a meta-analysis boils these studies' findings down to their essence. This makes even more imperative the need to ensure that meta-analyses assess these replications in valid, reliable and methodologically precise ways. As such, we recommend the use of the MMRQG instrument introduced and pilot-tested in this study on the material of 52 educational technology meta-analyses.

It is also true that meta-analyses should be kept up to date, especially in a rapidly changing field like educational technology, where implementation of new technologies and the pedagogies that drive them can be so expensive and difficult to reverse. One of the findings of this study which extends beyond this particular literature is that the MMRQG, or some similar instrument used by experienced and knowledgeable reviewers, can help discriminate between good and poor reporting (and by extension good and poor practices) and link this distinction to differences between the average effect size outcomes used by consumers and reviewers. It would be interesting to examine whether the inverse relationship found here between reporting quality and effect size generalises to other aspects of educational technology, or to education more broadly. Major organisational decisions regarding curriculum, pedagogy and related infrastructure are often informed by this research. To what extent are the results they rely upon biased?

The final word in this study of meta-analyses is that quality matters, certainly in terms of the primary research that goes into meta-analyses, but no less in the practices employed by meta-analysts themselves in synthesising, interpreting and reporting the results of their research efforts fairly and accurately.

### **Funding**

Parts of this project were supported by a grant from the Social Sciences and Humanities Research Council of Canada.

### **References**

(\* indicates reviews included in the analysis and cited in the text; for the complete list please contact us.)

- Abrami, P. C., & Bernard, R. M. (2012). Statistical control versus classification of study quality in meta-analysis. *Effective Education*, 4(1), 43–72. <https://doi.org/10.1080/19415532.2012.761889>
- Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research*, 82(4), 436–476. <https://doi.org/10.3102/0034654312458162>

- Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of Educational Technology & Society*, 11(4), 29–40. <https://www.learntechlib.org/p/75072/>
- Bernard, R. M. (2014). Things I have learned about meta-analysis since 1990: Reducing bias in search of “The Big Picture” / Ce que j’ai appris sur la méta-analyse depuis 1990 : réduire les partis pris en quête d’une vue d’ensemble. *CJLT: Canadian Journal of Educational Technology*, 40(3), 1–17. <https://doi.org/10.21432/T2MW29>
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79(3), 1243–1289. <https://doi.org/10.3102/0034654309333844>
- Bethel, E. C., & Bernard, R. M. (2010). Developments and trends in synthesizing diverse forms of evidence: Beyond comparisons between distance education and classroom instruction. *Distance Education*, 31(3), 231–256. <https://doi.org/10.1080/01587919.2010.513950>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. <http://https://doi.org/10.1002/9780470743386>
- Carpenter, C. R., & Greenhill, L. P. (1955). *An investigation of closed-circuit television for teaching university courses* (Report 1). Pennsylvania State University.
- Carpenter, C. R., & Greenhill, L. P. (1958). *An investigation of closed-circuit television for teaching university courses* (Report 2). Pennsylvania State University.
- \*Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7(3), 198–215. <https://doi.org/10.1016/j.edurev.2012.05.002>
- \*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. <https://doi.org/10.1016/j.edurev.2013.01.001>
- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Clark, R. E. (1985). Confounding in educational technology research. *Journal of Educational Computing Research*, 1(2), 137–148. <https://doi.org/10.2190/HC3L-G6YD-BAK9-EQB5>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Taylor & Francis. <https://doi.org/10.4324/9780203771587>
- Cook, D. A. (2009). The failure of e-learning research to inform educational practice, and what we can do about it. *Medical Teacher*, 31(2), 158–162. <https://doi.org/10.1080/01421590802691393>
- Cooper, H. M. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (6th ed.). Sage.
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (1994). *Handbook of research synthesis and meta-analysis*. Sage.
- Cuban, L., & Jandric, P. (2015). The dubious promise of educational technologies: Historical patterns and future challenges. *E-Learning and Digital Media*, 12(3-4), 425–439. <https://doi.org/10.1177/2042753015579978>
- \*D’Angelo, C., Rutstein, D., Harris, C., Bernard, R. M., Borokhovski, E., & Haertel, G. (2014). *Simulations for STEM learning: Systematic review and meta-analysis*. SRI International. <https://www.sri.com/publication/simulations-for-stem-learning-systematic-review-and-meta-analysis-full-report/>
- Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5(10), 3–8. <https://doi.org/10.3102/0013189X005010003>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press. <https://doi.org/10.1016/C2009-0-03396-0>
- Institute of Education Sciences. (n.d.). *What Works Clearinghouse*. <https://ies.ed.gov/ncee/wwc/>
- Jackson, G. B. (1980). Methods for integrated review. *Review of Educational Research*, 50(3), 438–460. <https://doi.org/10.3102/00346543050003438>
- Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A-M. K., Hammerstrøm, K. T., & Sathe, N. (2017). *Searching for studies: A guide to information retrieval for Campbell Systematic Reviews*. The Campbell Collaboration. [https://www.campbellcollaboration.org/images/Campbell\\_Methods\\_Guides\\_Information\\_Retrieval.pdf](https://www.campbellcollaboration.org/images/Campbell_Methods_Guides_Information_Retrieval.pdf)
- \*Kulik, C. L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior*, 7(1), 75–94. [https://doi.org/10.1016/0747-5632\(91\)90030-5](https://doi.org/10.1016/0747-5632(91)90030-5)

- Kulik, J. E., & Kulik, C.-L. C. (1987, February 26 – March 1). *Computer-based instruction: What 200 evaluations say* [Paper presentation]. Association for Educational Communications and Technology Annual Convention, Atlanta, GA, United States of America.
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad and ugly. *The Annals of the Academy of Political and Social Science*, 587(1), 69–81. <https://doi.org/10.1177/0002716202250791>
- Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis*. Sage.
- Littell, J. H., & White, H. (2018). The Campbell Collaboration: Providing better evidence for a better world. *Research on Social Practice*, 28(1), 6–12. <https://doi.org/10.1177/1049731517703748>
- Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of Web-based and lecture courses. *Journal of Experimental Psychology: Applied*, 8, 85–98. <https://doi.org/10.1037/1076-898x.8.2.85>
- McGreal, R. (1994). Comparison of the attitudes of learners taking audiographic teleconferencing courses in secondary schools in northern Ontario. *Interpersonal Computing and Technology Journal*, 2(4), 11–23. <https://www.learntechlib.org/p/79021/>
- Pickup, D. I., Bernard, R. M., Borokhovski, E., Wade, A. C., & Tamim, R. M. (2018). Systematically searching empirical literature in the social sciences: Results from two meta-analyses within the domain of education. *Russian Psychological Journal*, 15(4), 245–265. <https://doi.org/10.21702/rpj.2018.4.10>
- Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, 86(1), 207–236. <https://doi.org/10.3102/0034654315582067>
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Sage Publications.
- Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research*, 84(3), 328–364. <https://doi.org/10.3102/0034654313500826>
- \*Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., Wade, C. A., & Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers & Education*, 72, 271–291. <https://doi.org/10.1016/j.compedu.2013.11.002>
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage. <https://dx.doi.org/10.4135/9781483398105>
- Slavin, R. E. (1996). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15(9), 5–11 <https://doi.org/10.3102/0013189X015009005>
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, 81(1), 4–28. <https://doi.org/10.3102/0034654310393361>
- Valentine, J. C., & Cooper, H. M. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (Study DIAD). *Psychological Methods*, 13(2), 130–149. <https://doi.org/10.1037/1082-989X.13.2.130>
- Valentine, J. C., & Thompson, S. (2013). Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions. *Research Synthesis Methods*, 4(1), 26–35. <https://doi.org/10.1002/jrsm.1064>
- Valentine, J. C., Cooper, H. M., Patall, E. A., Tyson, D., & Robinson, J. C. (2010). A method for evaluating research syntheses: The quality, conclusions, and consensus of 12 syntheses of the effects of after-school programs. *Research Synthesis Methods*, 1(1), 20–38. <https://doi.org/10.1002/jrsm.3>
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <https://doi.org/10.1002/jrsm.11>

**Corresponding author:** Rana M. Tamim, [Rana.tamim@zu.ac.ae](mailto:Rana.tamim@zu.ac.ae)

**Copyright:** Articles published in the *Australasian Journal of Educational Technology* (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.

**Please cite as:** Tamim, R. M., Borokhovski, E., Bernard, R. M., Schmid, R. F., Abrami, P. C., & Pickup, D. I. (2021). A study of meta-analysis reporting quality in the large and expanding literature of educational technology. *Australasian Journal of Educational Technology*, 37(4), 100-115. <https://doi.org/10.14742/ajet.6322>