Computers & Education 72 (2014) 271–291

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The effects of technology use in postsecondary education: A meta-analysis of classroom applications

Richard F. Schmid [a,*], Robert M. Bernard [a], Eugene Borokhovski [a], Rana M. Tamim [b], Philip C. Abrami [a], Michael A. Surkes [a], C. Anne Wade [a], Jonathan Woods [a]

[a] Centre for the Study of Learning and Performance, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, QC, Canada H3G 1M8
[b] Zayed University, College of Education, P.O. Box 19282, Academic City, Dubai, United Arab Emirates

ARTICLE INFO

Article history:
Received 20 March 2013
Received in revised form
1 November 2013
Accepted 6 November 2013

Keywords:
Technology integration
Postsecondary education
Meta-analysis
Cognitive support tools

ABSTRACT

This meta-analysis is a study of the experimental literature of technology use in postsecondary education from 1990 up to 2010 exclusive of studies of online or distance education previously reviewed by Bernard et al. (2004). It reports the overall weighted average effects of technology use on achievement and attitude outcomes and explores moderator variables in an attempt to explain how technology treatments lead to positive or negative effects. Out of an initial pool of 11,957 study abstracts, 1105 were chosen for analysis, yielding 879 achievement and 181 attitude effect sizes after pre-experimental designs and studies with obvious methodological confounds were removed. The random effects weighted average effect size for achievement was $g+ = 0.27$, $k = 879$, $p < .05$, and for attitude outcomes it was $g+ = 0.20$, $k = 181$, $p < .05$. The collection of achievement outcomes was divided into two sub-collections, according to the amount of technology integration in the control condition. These were no technology in the control condition ($k = 479$) and some technology in the control condition ($k = 400$). Random effects multiple meta-regression analysis was run on each sub-collection revealing three significant predictors (subject matter, degree of difference in technology use between the treatment and the control and pedagogical uses of technology). The set of predictors for each sub-collection was both significant and homogeneous. Differences were found among the levels of all three moderators, but particularly in favor of cognitive support applications. There were no significant predictors for attitude outcomes.

1. Introduction

It must seem quite incredible to those outside educational circles that educators are still concerned about, even debating, the benefits of computers and associated computer technologies on teaching and learning in higher/postsecondary education. After all, computing is all-pervasive in general society and impacts significantly on every aspect of our daily lives, whether we are directly aware of it or not. But the utility of educational technology in general, and computing in particular, has been judged to be ancillary rather than fundamental (Clark, 1983, 1994). Why is this? It is partly because technology, in its earliest manifestations, was used to deliver content, and it was easily argued that any medium including live teachers could do that just as effectively. Media studies of the 1950s on distributed closed-circuit television versus live teaching (Carpenter & Greenhill, 1955, 1958) were among the first to demonstrate that there were no differences between live teachers and televised teachers. However, following Clark's logic, even the effects achieved by sophisticated technologies like those in medical training simulators (e.g., Haque & Srinivasan, 2006) could have been matched in less sophisticated ways. This argument effectively sidelines technology as inconsequential to the mission of postsecondary institutions, or at least relegates it to a secondary role. Of course there are those who have argued on the other side of the issue, usually citing the promise of technology to significantly affect learning and attitudes (Dede, 1996, 2004; Kozma, 1991, 1994; Mayer, 2008).

* Corresponding author. LB-579. Tel.: +1 514 848 2424x2001; fax: +1 514 848 4520.
E-mail addresses: schmid@education.concordia.ca (R.F. Schmid), bernard@education.concordia.ca (R.M. Bernard), eborokhovski@education.concordia.ca (E. Borokhovski), rana.tamim@zu.ac.ae (R.M. Tamim), abrami@education.concordia (P.C. Abrami), surkes@education.concordia.ca (M.A. Surkes), wada@education.concordia.ca (C.A. Wade), woodsjon@gmail.com (J. Woods).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2013.11.002

The crux of the contrary argument revolves not around the impact of technology per se, but how it is used (Bransford, Brown, & Cocking, 2000; Kozma, 2003). Like any tool, its effectiveness is determined by the purpose it serves and the manner in which it is implemented. As such, our definition of technology is not narrowly bound by issues of hardware and software, but instead essentially agrees with Ross, Morrison, and Lowther (2010) that educational technologies should not be considered unique entities in the learning process, but rather as “a broad variety of modalities, tools, and strategies for learning, [whose] effectiveness, therefore, depends on how well [they] help teachers and students achieve the desired instructional goals” (p. 19). According to this view, technology and pedagogy are in a synergistic relationship that is difficult to disentangle. Bernard et al. (2004) summarized evidence demonstrating this synergy between technology and pedagogy in distance education, and there is reason to believe that the same principle applies to postsecondary classroom applications.

One question that continually nags at Clark’s argument is whether more recent technologies, such as cognitive tools, communication tools and search and retrieval tools, are as benign as he has led us to believe about presentational tools. How do students learn through technology without being taught directly? One answer to this question, advanced by Cobb (1997), argues that the most compelling role of computing in learning is its ability to afford “cognitive efficiencies” to students. He says that in most if not all learning situations there is shared cognition among the learner, the task itself and the tools that the learner uses in the process. Further, he argues that the more learners can distribute cognition “outside of their heads,” the more cognition can be devoted to the process of learning new material. Cobb states: “The way forward in media design is to model learner and medium as distributed information systems, with principled, empirically determined distributions of information storage and processing over the course of learning” (pp. 32–33). In a similar vein, Mayer (2008) summarizes evidence-based principles of multimedia learning emphasizing the importance of dual channel encoding and the reductions in cognitive load when coherent and convergent content is presented to learners. Therefore, one purpose of the current review was to explore whether and to what extent cognitive tools promote student achievement and student attitudes toward instruction in learning environments involving technology, compared with other functions of technology use such as communication, search and retrieval and the presentation/reception of content.

Finally, the fact remains that decisions regarding the acquisition of and support for the use of technology in the schools too often remains woefully ill informed (Weston & Bain, 2010), based more on anecdotal evidence and personal convictions than on empirical research (e.g., the widespread implementation of 1:1 laptop programs, Murray & Olcese, 2011). While discussions about the use of media are plentiful (e.g., Clark & Mayer, 2011; Grabe & Grabe, 2006) and much research has been done, there appears to be no recent large-scale review that systematically examines the impact of contemporary technologies on postsecondary classroom learning exclusive of distance and online learning (see Bernard et al., 2004). The evidence provided by the current meta-analysis should prove important in informing instructional design and resource allocation decisions.

1.1. *Previous meta-analyses of technology integration*

Literally thousands of primary studies have been conducted comparing technology-enhanced conditions to regular classrooms without technology. Not surprisingly, meta-analysts have attempted to characterize this literature and estimate the effects of technology’s influence on learning. The most important reviews are listed in Table 1. The reviews cover the period 1995–2009, and more than half (i.e., 5 out of 8) examine computer technologies termed CAI or Computer-Assisted Learning. Zhao (2003) looks at Instructional Communication Technologies (ICTs) and two meta-analyses, Michko (2007) and Schmid et al. (2009), examine the breadth of technologies used in postsecondary education. As for subject matter, five address specific content areas. Interestingly, there are only two average effect sizes that diverge significantly from the others (Koufogiannakis & Wiebe, 2006; Zhao, 2003).

Collectively, these meta-analyses provide an interesting portrait of the state of instructional technology in postsecondary education over about a decade. Individually, however, none serves to truly summarize the state of research across time periods, technology types and subject matters.

The current meta-analysis examines the pedagogical uses of technology in classroom and blended learning, exclusive of distance and online learning, and attempts to bring together in broad strokes issues of the effects of teacher and learner use of a range of technologies currently in service of achievement and attitude outcomes.

With regard to the design of technology studies, Tamim, Bernard, Borokhovski, Abrami, and Schmid (2011), Bernard et al. (2009) and Cook (2009), among others, have argued that research that pits a technology in the experimental condition against a no-technology control condition, is simply a proof of concept study (i.e., Does technology work?) and does little to help us understand how to use technology. Needed are nuanced studies comparing a technology to another technology or some enhancement of that technology (i.e., Does this technology work better than that one, and why?). This meta-analysis includes studies of the with or without type along with the growing corpus of studies that compares one technology condition against another technology condition, answering these more nuanced questions. One of the purposes, then, is to investigate core instructional questions and to determine if, and if so, how the answers differ across these two design patterns.

Table 1
Recent meta-analyses of technology use in postsecondary education and their characteristics.

Meta-analyses	k	ES+[a]	Type of technology	Subject matter
Fletcher-Flinn and Gravatt (1995)	120	0.24	CAI	Combination
Hsu (2003)	25	0.43	CAI	Mathematics
Zhao (2003)	9	1.12	ICT	Language
Koufogiannakis and Wiebe (2006)	8	−0.09	CAI	Information literacy
Timmerman and Kruepke (2006)	5	0.24	CAI	Combination
Michko (2007)	45	0.43	All	Engineering
Schenker (2007)	46	0.24	CAI	Mathematics
Schmid et al. (2009)	310	0.28	All	Combination

[a] ES+ means average effect size.

1.2. Answering the "big question": does technology work?

In 2011, Tamim et al. published a second-order meta-analysis of 25 meta-analyses that cut across all levels of formal education, subject matters and technology types, from the 1970s to the present. They found a weighted average effect size of $+0.35$ ($p < .01$) encompassing 1055 primary studies and 109,700 participants. Tamim et al. answered the big question based on the kinds of studies and meta-analyses that tested the with and without question. They concluded that technology does enhance learning, even if only to a relatively small degree. But as Cooper and Koenka (2012) point out, a second-order meta-analysis remains a limited means of addressing the host of issues that can only be settled in a first-order meta-analysis, where coding decisions are made by the meta-analyst and the synthesis is conducted at the more granular level of the individual effect sizes extracted from primary studies. It is these issues, plus the fact that the educational research landscape has changed, to largely include technology vs. technology types of instructional settings, that motivated our effort to perform a primary meta-analysis of technology use in postsecondary education in the modern classroom.

1.3. Technology in the postsecondary classroom

There are numerous, important differences between K-12 and postsecondary education. The following differences are self-evident: the former is mandatory, the latter voluntary (e.g., self-motivated); exposure time in K-12 is usually greater; prior knowledge, a key factor in learning, is both significantly greater as education progresses, and qualitatively different; and the nature/complexity of the content matter increases significantly. Finally, theory shows that cognitive, affective and psychomotor development from early childhood through adulthood calls for differing pedagogical strategies to support qualitatively different types of learning and motivation (O'Donnell, D'Amico, Schmid, Reeve, & Smith, 2008). The postsecondary teaching and research community will benefit from an empirical examination of what works vis-à-vis technology integration. In order to capture contemporary technologies, this analysis is restricted to studies published from 1990 up to 2010.

To summarize, the current meta-analysis is unique in the following ways: 1) it is the most comprehensive meta-analysis of technology use in postsecondary classroom education ever conducted; 2) it looks at the pedagogical uses of a broad range of educational technologies; 3) it examines important instructional moderator variables; 4) it looks at both achievement and attitudes of postsecondary students; and 5) it is refined because it analyses separately two major sub-collections of studies: no technology in the control condition and some technology in the control condition.

1.4. Research questions

This meta-analysis was designed to answer questions about the impact of instructional technology on postsecondary students' achievement and attitudes as both a combined collection of studies, and as two sub-collections of studies: 1) no technology in the control condition; and 2) some technology in the control condition.

In addition, it looked at a set of moderator variables: 1) levels of education; 2) subject matter; 3) classroom/blended learning; 4) difference between treatment and control in technology use; and 5) pedagogical uses of technology. The specific research questions follow.

1. What is the weighted average effect size and variability for studies that investigate the impact of higher degrees of the instructional uses of computer-based technology (as it compares to lower degrees of instructional uses of computer-based technology) on postsecondary students' achievement and attitudes?
2. What is the weighted average effect size and variability for the two separate sub-collections, identified above, in terms of achievement and attitude outcomes?
3. For studies with no technology in the control condition, what moderator variables significantly predict achievement and attitude outcomes?
4. For studies with some technology in the control condition, what moderator variables significantly predict achievement and attitude outcomes?
5. Has the effectiveness of instructional technology changed over the twenty years covered by this study?

2. Method

2.1. Definition of terms

The following definitions guided this review as they informed the design of inclusion/exclusion criteria and decisions for coding study features.

- Educational technology is understood as Ross et al. (2010) describe, as "a broad variety of modalities, tools, and strategies for learning, [whose] effectiveness, therefore, depends on how well [they] help teachers and students achieve the desired instructional goals" (p. 19), and includes all types of computer-based tools and applications.
- The difference in the degree of technology use describes the distinction between the experimental and control conditions. The effect of educational use of technology for teaching and learning purposes is determined by comparing higher technology use in the experimental condition to lower technology use (including no technology use) in the control condition.
- Formal postsecondary educational settings are those that offer instruction for credit leading to a degree, diploma or any other form of academic advancement.
- Blended learning occurs when at least 50% of the instructional activity is performed in a classroom and a substantial amount is conducted online with either individual students or groups of students. Distance education and fully online courses, where there is little or no face-to-face meetings of the class, were excluded from this study.

- Pedagogical uses of technology were coded as the following major functions:
  1) To promote communication and/or facilitate the exchange of information. This category includes technology that enables a higher level of interaction between individuals (i.e., two-way communications among learners and between learners and the teacher);
  2) To provide cognitive support for learners. This category encompasses various technologies that enable, facilitate, and support learning by providing cognitive tools (e.g., concept maps, simulations, wikis, different forms of elaborate feedback, spreadsheets, word processing). These technological tools and applications are primarily for student use;
  3) To facilitate information search and retrieval. This type of technology is intended to enable and/or facilitate access to additional information (e.g., web-links, search engines, electronic databases); and
  4) To enable or enhance content presentation. Technology in this category is primarily used by teachers to present or deliver, illustrate and otherwise enrich the content of instruction (e.g., PowerPoint presentations, graphical visualizations, computer tutorials with limited interactive features).

When more than one purpose was identified, codes indicating multiple purposes were created (e.g., cognitive support plus presentational support). For additional information, see the Codebook in Appendix A.

- Achievement is any objective measure of academic performance (e.g., exams or test scores) including but not limited to standardized and validated measures.
- Student attitudes toward the learning environment includes student ratings of their learning experience (e.g., satisfaction with the course and teacher, self-evaluation in terms of learning progress).
- Difference in degree of technology use between the treatment and control is described as being either low, moderate, or high and may be attributed to one (or a combination of) the following sources:
  1) More intensive use of the same technology (i.e., more frequently and/or for a longer time period);
  2) Use of more advanced technologies (i.e., additional functions/features in similar or different technologies); and
  3) Use of a larger versus smaller number of different technological tools, devices, software, etc.
- Pedagogical moderator variables are characteristics that are largely under the control of instructors, such as the purpose of technology use, the degree of blended instruction, and the extent of technology use. By contrast, contextual moderator variables are those characteristics that are largely not under the control of instructors, such as course duration, subject matter, and students' educational level.

2.2. Literature search strategies and data sources

Extensive literature searches were designed to identify and retrieve primary empirical studies relevant to the research questions. Key terms used in search strategies, with some variations to account for specific retrieval sources, included: ("technolog*," "comput*" "web-based instruction," "online," "Internet," "blended learning," "hybrid course*," "simulation," "electronic," "multimedia" OR "PDAs" etc.) AND ("college*," "university," "higher education," "postsecondary," "continuing education," OR "adult learn*") AND ("learn*" "achievement*," "attitude*," "satisfaction," "perception*," OR "motivation," etc.), but excluding "distance education" or "distance learning" in the subject field. To review the original search strategies and see the complete search results please visit [http://doe.concordia.ca/cslp/cslp_cms/?q=node/424].

The following electronic databases were among the sources examined: ERIC (WebSpirs), ABI InformGlobal (ProQuest), Academic Search Premier (EBSCO), CBCA Education (ProQuest), Communication Abstracts (CSA), EdLib, Education Abstracts (WilsonLine), Education: A SAGE Full-text Collection, Francis (CSA), Medline (PubMed), ProQuest Dissertation & Theses, PsycINFO (EBSCO), Australian Policy Online, British Education Index, and Social Science Information Gateway.

In addition, Google Internet searches were performed to help identify gray literature, including a search for conference proceedings. Review articles and previous meta-analyses were used for branching, and the tables of contents of major journals in the field of educational technology (e.g., *Educational Technology Research & Development*) were manually searched.

Summary tables of general and specific publication sources appear in Appendix B, Tables B1 and B2.

2.3. Inclusion/exclusion criteria and review procedure

This review employed a liberal approach to study inclusion. Instead of excluding studies of questionable methodological quality *a priori*, the approach used involved coding for research design and other methodological study features to enable their subsequent analysis (Abrami & Bernard, 2013).

Recognizing that the role of computer-based technologies in education has changed over time in parallel with the rapid developments of the tools themselves, a decision regarding the start date for the review had to be made. Since the early 1980s, the classroom use of computers has escalated dramatically, but the year 1990 in particular is generally regarded as the date when Internet access with personal computers became widely available for educational purposes (e.g., Kozma, 1994), and computers had both substantially increased processing capacity and advanced functions (e.g., the introduction of the Macintosh Classic in 1990). Thus, the decision was made to consider 1990 the start date for the review.

The procedure for selecting studies involved two stages. First, studies identified through literature searches were screened at the abstract level. Second, the review of full-text documents was carried out. To be included a study had to have the following characteristics:

- be published no earlier than 1990;
- be publicly available (or archived);
- address the impact of computer technology on students' achievement and/or attitudes;
- contain at least one between-group comparison where one group is considered the experimental condition and the other group the control condition, using the criterion of the degree of difference of technology use (from none to high);

- be conducted in a formal postsecondary educational setting;
- represent classroom or blended instruction but not distance education environments; and
- contain sufficient statistical information for effect size extraction.

Failure to meet any of these criteria led to the exclusion of the study with the reason(s) for rejection documented for future reporting. Two researchers working independently rated studies, first at the abstract level, then at the full text level, on a scale from 1 (definite exclusion) to 5 (definite inclusion). All disagreements were discussed until they were resolved and initial agreement rates were calculated as Cohen's Kappa (κ) and as Pearson's *r*. Similarly, two researchers participated in all other data extraction procedures (i.e., effect size calculation and study features coding). Agreement rates for their independent decisions at these stages of the review are reported as Cohen's Kappa (κ) and can be found at the beginning of the Results.

2.4. Experimental and control conditions

2.4.1. Designation of treatment and control

The degree of technology use was the primary determinant for assigning groups to either the experimental or the control condition. This distinction is important because it specifies the $\pm$ valence of the effect size. Two types of studies were found, those that contained no technology in the control condition and those that contained some technology in the control condition. In the former class of studies, the assignment of the experimental and control group designation was clear. In the latter case, the differential use of technology in the two conditions was rated. The condition containing the highest degree of technology use was designated the treatment condition and the alternative condition was the control. There were three possible interpretations of the degree of technology use. The experimental group was considered: a) to contain the longest or most frequent exposure to technology tools; b) to contain more advanced technologies (i.e., the ones that enable more functions/features); and/or c) to employ a larger number of technology tools. If the two conditions were judged to be equal on all of the above criteria, the study was rejected. In this sense, degree of technology use served as an additional inclusion/exclusion criterion.

Examples. Following are examples of equal degree of technology use whereby the studies were rejected. Katayama and Crooks (2001) investigated the effect of the type of notes (messages to students) designed to help students in three different experimental conditions to complete their computer assignments. Notes were different in their content and structure, but all were delivered by means of the same technology in compatible amounts. In a study by Jackson (2002), the only difference between the conditions was a single additional e-mail message sent by the teacher to students in one group to boost their self-efficacy. Ford and Chen (2000) studied the effect on learning outcomes on pedagogical approaches that take into account individual learning styles. Subsequently, the manipulated experimental factor in this study was matching versus not matching students' learning styles, achieved without any difference in technology use.

2.4.2. Degree of difference between the treatment and control

The same criteria noted above were used to determine the difference in degree of technology use, but in addition, the determination of the magnitude of this difference was made. Experimental and control conditions were evaluated by independent reviewers, first individually and then in comparison with each other to reach consensus resulting in one of the following conclusions.

- The difference between experimental and control conditions is low. This code applied when: a) treatment and control groups both used the same technology, but one group spent slightly more time on it than the other, or b) one group engaged in an intervention that employed a very small amount of technology use while the other group did not use technology at all, or c) the treatment group used technology that provided some minor extra features as compared to the control group.

Examples. In Pemberton, Borrego, and Cohen (2006), one group used computer-aided instruction (CAI) only to review course materials on-line, the other group reviewed the materials face-to-face. In Yu and Yu (2002) one group received and submitted assignments by e-mail, while the other exchanged paper-based assignments. In Huang (1995) CAI with elaborated feedback was compared with CAI with verification feedback (correct/incorrect response).

- The difference between experimental and control conditions is moderate. This code was used to describe studies where: a) one group was provided with more than a minimal technology use combined with other instructional methods and the other group worked without technology, or b) when both groups used technology but one group spent much more time on it than the other (e.g., close to twice as much).

Examples. In a study by Chen and Levinson (2006), one group used simulation software for all homework assignments (but not in class), while the other group did not. In Ford and Klicka (1994), supplemental use of CAI was required by the course design, whereas for the control group it was optional.

- The difference between experimental and control conditions is high. The application of this code was warranted when one group worked exclusively with technology throughout the entire intervention and the other group worked without technology or when the complexity, sophistication or diversity of technological tools used in one group was substantially higher than in the other group.

Examples. In a study by Barnes (1994), a multimedia laboratory condition was compared to no multimedia. In Behnke and Ghiselli (2004), full-scale computer-based instruction was compared with traditional (technology-free) lecture.

Coders also made decisions about the nature of the achievement and attitude outcomes and the relevance and sufficiency of data for effect size extraction.

2.5. Study feature coding

Coded moderator variables (i.e., study features) were used to explore between-study variability in effect sizes. The study features were largely based on those used by Bernard et al. (2009) in studying distance education, but were modified as needed to attend to unique classroom-based and blended learning issues. Study features included publication variables (e.g., date of publication), methodological quality variables (e.g., research design), pedagogical uses of technology (e.g., cognitive support, presentation support, multiple purposes) along with other instructional and contextual moderator variables.

A methodological quality index (see Abrami & Bernard, 2013), reflecting the rigor and reliability of various components of primary studies design and implementation (Valentine & Cooper, 2008), as well as precision of basic meta-analytical procedures, included seven individually coded study features, as follows: 1) primary study research design; 2) equivalence of learning materials; 3) instructor equivalence; 4) treatment duration; 5) psychometric quality and type of the outcome assessment tool; 6) effect size extraction method; and 7) publication type (including the level of impact factor for studies published in peer-reviewed journals). The complete codebook is available in Appendix A.

2.6. Effect size calculation and synthesis

2.6.1. Calculating effect sizes

A *d*-type standardized mean difference effect size was used as the common metric (i.e., Cohen's *d*). This equation is expressed as ($d = \overline{X}_E - \overline{X}_C/SD_{Pooled}$). A modification of this basic equation was used for studies reporting pretest and posttest data for both experimental and control groups ($d_{gain} = \overline{D}_E - \overline{D}_C/SD_{Posttest/Pooled}$). When descriptive statistics were not available, effect sizes were extracted from inferential statistics, such as *t*-tests, *F*-tests, or exact *p*-values, using conversion equations from Glass, McGaw, and Smith (1981) and Hedges, Shymansky, and Woodworth (1989). In some cases, authors reported effect sizes in addition to other statistical information. These effect sizes were recorded and tested against effect sizes calculated from complete descriptive data or inferential statistics. When an author's estimated effect size could not be verified independently, the study was dropped. To correct for small sample bias, *d* was converted to the unbiased estimator *g* (Hedges & Olkin, 1985).

2.6.2. Synthesizing effect sizes

The random effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hedges & Olkin, 1985) was judged to be the correct analytical approach for this meta-analysis. In the random effects model, effect sizes are weighted by the inverse of the sum of their within-study variance (V_i) and average between-study variance (tau-squared). As a result there is no between-study variance left unaccounted for after the analysis is performed. For a full explanation of this, see Borenstein, Hedges, Higgins, and Rothstein (2010). The random effects model was used to interpret and report average effect sizes ($g+$), standard errors (SE), confidence intervals (lower 95th and upper 95th) and *z*-values with associated *p*-values.

The fixed effect model, where effect sizes are weighted only by the inverse of their within-study variance (V_i), was used to estimate total between-study variability (Q_{Total}) and test for heterogeneity. I^2 (i.e., percentage of heterogeneity in effect sizes exceeding chance sampling expectations, Higgins, Thompson, Deeks, & Altman, 2003) is reported and interpreted.

Moderator variable analysis was conducted in three phases. First, for each sub-collection, random effects meta-regression (Hedges & Olkin, 1985; Pigott, 2012) was used to create models of the set of moderating variables previously identified. Since all of the moderator variables were categorical, dummy codes were created and used in hierarchical multiple regression. The complete models were reduced to a smaller set of variables based on the significance of their individual contribution ($Q_{Regression}$, *p*-value). Homogeneity of the reduced model was judged by the significance of the residual value ($Q_{Residual}$, *p*-value) on the last step.

Second, the significant moderators identified in regression analysis were then analyzed using the mixed effects model that contains elements of both the fixed effects and the random effects approaches. Average effect sizes for categories of the moderator were calculated using a random effects model. The variance component $Q_{Between}$ was calculated across categories using a fixed effect model (Borenstein et al., 2009). All analyses, including sensitivity and publication bias analysis, were performed in *Comprehensive Meta-Analysis*™ 2.2.048 (Borenstein, Hedges, Higgins, & Rothstein, 2008). Lastly, post hoc analyses (Bonferroni correction) were conducted on selected levels of the moderator variables.

3. Results

3.1. Total collection

Overall, 11,957 abstracts were reviewed, resulting in the retrieval of 3846 full-text studies potentially suitable for the analysis. Through a review of full-text documents, 1105 studies (sources of included studies are described in Appendix B) were retained for further analysis (a summary of reasons for exclusion are in Appendix C). The included studies yielded 1669 effect sizes, 1376 of which were in the achievement category and 293 were in the attitude category.

3.2. Inter-rater reliability

At least two trained raters were involved in all stages of the review. Agreement rates pertinent to each stage were:

- Abstract screening – 86.89% (Cohen's $\kappa = 0.74$) or $r = 0.75$, $p < .001$.
- Full-text manuscript inclusion/exclusion – 85.57% ($\kappa = 0.72$) or $r = 0.84$, $p < .001$.
- Effect size comparison decisions (defining experimental and control condition, deciding on the number of effects and which data sources to use) – 90.03% ($\kappa = 0.80$).
- Effect size calculation (i.e., accuracy of data extraction and selection and application of equations) – 97.26% ($\kappa = 0.95$).
- Study features coding – 91.77% ($\kappa = 0.84$).

3.3. Achievement outcomes

3.3.1. Research design and methodological confounds

The questionable methodological quality of primary research studies is a principal source of confounding in meta-analyses (Lipsey, 2003). Efforts were made to neutralize as many of these as possible. The first step was to remove studies with obvious flaws in research design, especially pre-experimental designs where selection bias was likely to be present (Campbell & Stanley, 1963). Within the 1376 achievement effect sizes, 416 with pre-experimental designs were removed. Also removed were studies containing effect sizes reported by authors that could not be verified through calculation ($k = 6$), effect sizes where content-based materials had been supplied to the two groups differentially ($k = 37$) and studies of unknown or atypically long treatment duration ($k = 31$). Also removed were seven ($k = 7$) outlying effect sizes that ranged from $g = +2.62$ to $+8.95$, producing a more conservative estimate of average effect and reducing variability. The final collection contained 879 effect sizes derived from 674 studies ($N = 58{,}585$ students). In addition, studies that had been coded for an additional seven methodological qualities (see the Method, Section 2.5), based in part on the study DIAD (Valentine & Cooper, 2008), were summed into a weighted methodological index (see Abrami & Bernard, 2013) with values ranging from 0.0 to 14.0. Simple meta-regression analysis (random model) was used to test the relationship between the index values and effect sizes. Results indicated no relationship ($b_Y = -0.003$, $p = .75$, $Q_{Regression} = 0.10$, $p = .75$), suggesting that the effects of questionable research design and other potentially confounding methodological practices had been removed or neutralized.

3.3.2. Publication bias

Analysis of publication bias seeks to determine whether a sizable number of studies might have been missed or not included in a meta-analysis (Rothstein, Sutton, & Borenstein, 2005) and if this number would nullify the average effect (Classical Fail-Safe Analysis) or bring it to a trivial effect size magnitude (Orwin's Fail-Safe Analysis). There are sub-routines within the publication bias module of Comprehensive Meta-Analysis™ (Borenstein, Hedges, Higgins, & Rothstein, 2005) that allow for exploration of publication bias. For the achievement data ($k = 879$), an additional 197,712 null-effect sizes would be necessary to bring the overall probability to $\alpha = .05$. Additionally, 1423 null-effect sizes would be required to bring $g+$ to the trivial level of 0.10. Further examination of a funnel plot based on the random effects model, revealed near symmetry with no imputations required. Thus, we found a negligible likelihood of publication bias in the achievement outcomes. The same analyses were performed on attitude outcomes with a similar result.

3.3.3. Sensitivity analysis

Sensitivity analysis seeks to determine if there are any effect sizes that exert an undue influence on the central tendency and variability of the collection, often resulting from highly leveraged studies residing at the extremities of the distribution. When the final collection of 879 effect sizes was examined using the one study removed procedure in Comprehensive Meta-Analysis™ (Borenstein et al., 2005), no anomalies were found. The same was true for attitude outcomes.

3.3.4. Final collection of achievement data

Fig. 1 shows a histogram of the final collection of achievement 879 effect sizes. The distribution is nearly normal, not overly skewed (skewness = 0.28) but slightly leptokurtic (kurtosis = 1.21). The unweighted mean is 0.28 (very close to the random effects mean of 0.27) with a standard deviation of 0.56. The median is 0.25.

Table 2 presents statistics related to the final collection: the weighted average effect size ($g+$), number of effect sizes (k) and the lower and upper limits of the 95th confidence interval, all based on the random effects model. While $g+ = 0.27$ is statistically significant ($p < .05$), according to Cohen (1988) this is a small average effect size. In terms of difference in percentile rank (based on U_3, Cohen, 1988), 10.7% of students in the technology treatment condition would be expected to exceed students at the mean/median of the control condition.

The heterogeneity statistics derived from the fixed effect model appear in the final line of Table 2 and portray the collection as significantly heterogeneous with an I^2-value of 72.42%. According to Higgins et al. (2003) this is a moderately high to high degree of heterogeneity.

One of the potential confounds in this meta-analysis is the differential nature of the control conditions of studies with no technology in the control condition and those with some technology in the control condition. The designation of the treatment and control in these sub-collections was determined in two somewhat different ways as described in the Method section. Because of this difference, these two datasets were analyzed separately. Tables 3 and 4 show the statistical outcomes from the two sub-collections. The weighted averages in each sub-collection are significantly different from zero, and each is significantly heterogeneous and displays very similar I^2s (i.e., percentage of true heterogeneity). However, the sub-collection of some technology in the control condition has a somewhat larger average weighted effect size than the sub-collection with no technology in the control condition.

3.3.5. Moderator variable analysis

Moderator variable analysis seeks to establish relationships between coded study features and effect sizes so as to extend understanding of the average effect size. Meta-analysis and by extension moderator variable analysis, remains a descriptive observational methodology that cannot establish causal relationships, even when all of the studies it contains are randomized control trials (Borenstein et al., 2009). This is a very important consideration to bear in mind when reading this or any meta-analysis.

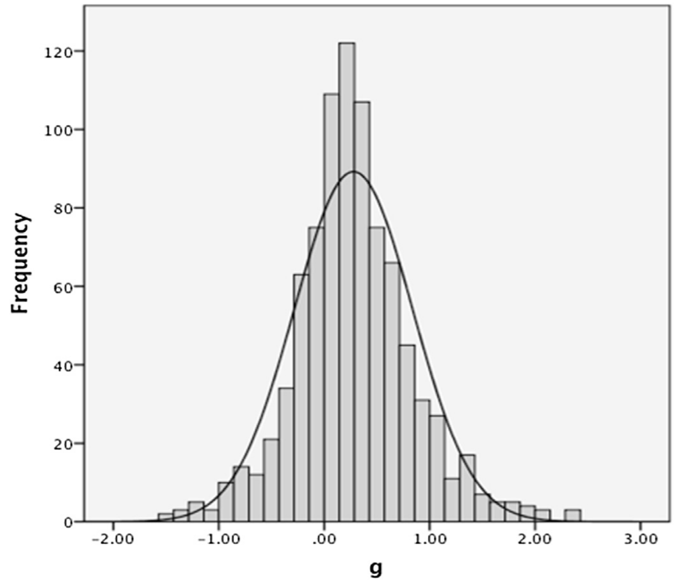

Fig. 1. Histogram of 879 unweighted effect sizes based on achievement outcomes.

According to Pigott (2012), the chances of moderator variables being confounded, so as to complicate interpretation, are fairly high in a meta-analysis, unlike primary experimental research where the researcher has more control over moderator variables and can introduce covariates to control for certain deficits in design. By contrast, analysts have control over little more than identification and coding of extant moderators.

A number of substantive moderators were identified and coded as described in the Method section. These were: 1) student demographics (i.e., level of education); 2) content area (i.e., STEM and non-STEM subjects); 3) instructional setting (e.g., conditions of classroom/blended learning); and 4) two technology conditions (i.e., degree of difference between treatment and control and the pedagogical uses of technology). Since the potential exists for two or more of these variables to be confounded, we tested these possibilities by examining cross-tabulations of the variables and multi-way interactions (i.e., confounds) using the meta-analytic equivalent of two-way *ANOVA* (Pigott, 2012). While there were some irregularities that might suggest variable overlap, there were no statistically significant interactions, suggesting that each variable could be interpreted independently with reasonable certainty.

Instead of treating each moderator variable separately, as is often necessary in meta-analyses containing many fewer studies than this one, we used multiple meta-regression as a way to assess the joint influence of a set of predictors (i.e., moderator variables) on achievement outcomes. Fig. 2 shows the analytical plan and the predictors that were entered at each step in the multiple regression models. In both sub-collections of studies, predictors related to demographics (i.e., level of education) and instructional context (i.e., conditions of classroom/blended learning) were not significant and were dropped from the final model, shown in Table 5 for no technology and Table 9 for some technology. However, the categorical demographic moderator variable, subject matter, was significant. That left subject matter, degree of difference between the treatment and control, and pedagogical uses of technology in the reduced model. As these were all categorical predictors, dummy codes were used to represent them in the model.

Multiple (hierarchical) meta-regression was performed in SPSS 23.0 using random effects inverse variance weights that are composed of two sources of variance: within-study variance (V_i) and average between-study variance (τ^2) (Hedges & Olkin, 1985; Pigott, 2012).

Note that most of the standard features of regression analysis appear in Tables 5 and 9. The slopes are accurate when meta-regression is run in statistical packages designed for primary studies. However, the standard errors are not and must be adjusted using the equation $\mathrm{SE}_{(\mathrm{Adjusted})} = \mathrm{SE}_{(\mathrm{Unadjusted})}/\sqrt{\mathrm{MS}_{(\mathrm{Residual})}}$. All other statistics using the standard error must also be adjusted. For the purposes of meta-analysis, R^2 is incorrect and cannot be interpreted (Pigott, 2012). The question becomes then, is the reduced set of predictors significant and does it result in homogeneity of effect size, indicating that the predictors are a reasonable fit to the data.

3.3.6. Moderator variable analysis: no technology in the control

As can be seen from the last lines of Table 5, both overall ES significance and homogeneity were found. Judging by the *Q*-values in regression (far right column shows the accumulated $Q_{\mathrm{Regression}}$), the variable with greatest influence in the model is degree of difference between the treatment and the control. The variable with the least influence is subject matter and pedagogical uses of technology are in-between. Since all the moderators were significant, all can be interpreted.

Table 2
Overall weighted average effect size (random effects model) for the final collection based on achievement and heterogeneity statistics (fixed effect model).

Population estimates	*k*	*g*+	SE	Lower 95th	Upper 95th
Final collection	879	0.27*	0.02	0.24	0.31
Heterogeneity analysis (fixed effect model)	$Q_T = 3183.10$ (df = 878), $p < .001$, $I^2 = 72.42$				

*$p < .01$.

Table 3
Overall weighted average effect size (random effects model) for studies with no technology in the control group on achievement and heterogeneity statistics (fixed effect).

Population estimates	k	$g+$	SE	Lower 95th	Upper 95th
Sub-collection no technology in the control	479	0.25*	0.02	0.21	0.29
Heterogeneity analysis	$Q_T = 1690.49$ (df = 478), $p < .001$, $I^2 = 71.72$, $\tau^2 = 0.15$.				

*$z = 22.59$, $p < .01$.

Table 4
Overall weighted average effect size (random effects model) for the studies with some technology in the control condition on achievement and heterogeneity statistics (fixed effect).

Population estimates	k	$g+$	SE	Lower 95th	Upper 95th
Sub-collection some technology in the control	400	0.31*	0.03	0.25	0.36
Heterogeneity analysis	$Q_T = 147.91$ (df = 399), $p < .001$, $I^2 = 73.18$, $\tau^2 = 0.20$.				

*$z = 21.16$, $p < .01$.

A second step in the analysis of was to perform an *ANOVA*-equivalent mixed effects moderator variable analysis on each variable remaining in the reduced models. This allows for the comparison of the predicted weighted means (first column in Tables 6–8) with the empirical weighted means (second column in Tables 6–8) and the assessment of the relative effects of the levels of these variables. In some cases post hoc analyses (Hedges & Olkin, 1985) were performed between levels or combinations of levels in order to further illuminate the results.

According to these analyses, STEM subject matters (Table 6) profit more by the use of technology versus no technology than non-STEM subjects. One of the most interesting features of these analyses (Table 7) is the large difference between both the low and medium technology use average effect sizes and the high use average effect size. Post hoc analysis revealed that low and medium use combined was significantly larger than high use ($z = +6.367$, $p < .01$) suggesting that there may be an upper threshold for using different amounts of technology to promote learning outcomes.

The results shown in Table 8 are somewhat more complicated, primarily because there are categories that include combinations of uses in difficult to specify amounts. One thing that is very clear from Table 8, CS and combos with CS are higher than all other categories except search and retrieval, which has only four studies (i.e., the effect size is also not significant). Not all of these categories are significantly different from one another, but the primary comparison between PS and CS is ($z = 3.68$, $p < .001$). Interestingly, the combos with CS and combos with CS + PS are significantly different from one another ($z = 3.29$, $p < .001$).

3.3.7. Moderator variable analysis: some technology in the control

There were 400 effect sizes with variable amounts of technology in both the treatment and the control groups. The results of the multiple meta-regression are shown in Table 9, and like the previous analysis the overall set of predictors is both significant and homogeneous.

One striking difference here is that subject matter is the most influential predictor, and the relationship between STEM and non-STEM subject matters is reversed from the first analysis. The effects of technology use are higher in subject matters such as the humanities, education, language, etc. (i.e., negative sign in Table 9). This counterintuitive result, shown more explicitly in Table 10, may be related to the

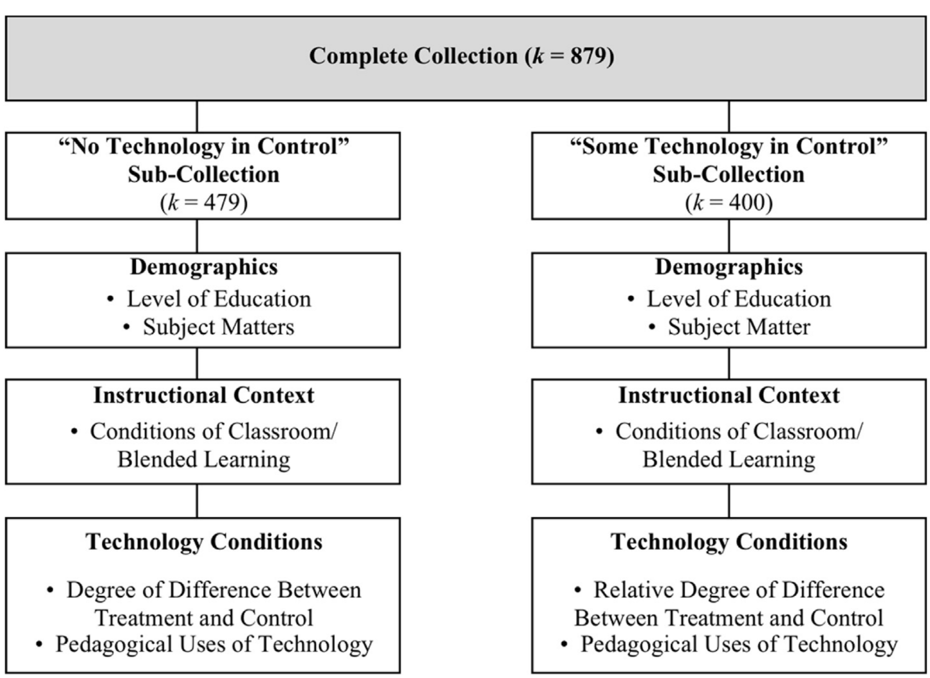

Fig. 2. Analytical models for multiple meta-regression for each of the sub-collections.

Table 5
Reduced meta-regression model for studies with no technology in the control condition (random model).

Variables in the final model	Slope (b_Y)	SE(adj)	Lower 95th(adj)	Upper 95th(adj)	z-Value(adj)	p-Value	$Q_{Regression}$
Intercept	0.48	0.15	0.19	0.77	3.27	.001	
Demographic – subject matter (reference category – Non-STEM)							
D-1 STEM	0.08	0.05	−0.01	0.17	1.89	.06	5.87*
Technology A – degree of difference between the treatment and the control (reference category – high)							
D1-Low	0.19	0.08	0.05	0.34	2.56	.01	
D2-Medium	0.24	0.05	0.14	0.34	4.84	.00	35.57**
Technology B – pedagogical uses of technology (reference category – all other combos w/o CS)							
D-1 Communication support	−0.06	0.17	−0.39	0.27	−0.36	.72	
D-2 Search & retrieval support	0.42	0.26	−0.09	0.93	1.63	.10	
D-3 Presentation support (PS)	0.00	0.14	−0.27	0.27	−0.00	1.00	
D-4 Cognitive support (CS)	0.16	0.14	−0.12	0.43	1.12	.26	
D-5 Combination (CS + PS)	−0.02	0.13	−0.27	0.23	−0.18	.86	
D-6 Combinations with CS	0.28	0.15	−0.01	0.570	1.88	.06	55.53***
$Q_{Residual}$							491.97****

*First step, df = 1, p = .02; **second step, df = 3, $p < .0001$; ***last step, df = 9, $p < .0001$; ****residual, df = 469, p = .22.

Table 6
Levels of subject matter (mixed effects model).

Levels	Effect sizes			Standard error and 95th CI			Q-value
	g+(pred)	g+(unadj)	k	SE	Lower 95th	Upper 95th	
STEM	0.31	0.32	169	0.03	0.26	0.38	
Non-STEM	0.24	0.21	310	0.03	0.15	0.267	
Between							6.44*

*df = 1, p = .011.

Table 7
Levels of degree of difference between the treatment and the control (mixed effects model).

Levels	Effect sizes			Standard error and 95th CI			Q-value
	g+(pred)	g+(unadj)	k	SE	Lower 95th	Upper 95th	
Low	0.28	0.29	57	0.05	0.19	0.38	
Medium	0.34	0.34	261	0.03	0.29	0.40	
High	0.07	0.07	161	0.04	−0.003	0.14	
Between							34.62*

*df = 2, $p < .001$.

nature of the some technology research design. Adding something else (i.e., the treatment condition) to an already existing technology may be more important to the success of non-STEM students than it is to STEM students, where the presence of technology is more likely to be standard practice (e.g., simulations, virtual laboratories).

Contrary to the previous analysis of no technology studies, the moderator variable relating to the degree of difference between treatment and control contributes only modestly to the model. The signs are positive, indicating that low and medium outperformed high, but $Q_{Regression}$ is considerably diminished by comparison. Similar to the previous analysis, the contribution by pedagogical uses is modest. Likewise, cognitive support stands out as a nearly significant predictor compared to the reference category.

Table 8
Levels of pedagogical uses of technology (mixed effects model).

Levels	Effect sizes			Standard error and 95th CI			Q-value
	g+(pred)	g+(unadj)	k	SE	Lower 95th	Upper 95th	
Communication	0.20	0.22	17	0.07	0.09	0.36	
Search & retrieval	0.70	0.72	4	0.37	0.001	1.43	
Presentation (PS)	0.17	0.17	65	0.05	0.06	0.27	
Cognitive (CS)	0.40	0.41	56	0.07	0.26	0.56	
Combination (CS + PS)	0.20	0.20	293	0.03	0.15	0.25	
Combinations with CS	0.53	0.53	30	0.10	0.34	0.73	
Combinations w/o CS	0.21	0.21	14	0.13	−0.05	0.48	
Between							19.61*

*df = 6; p = .003.

Table 9
Reduced meta-regression model for studies with some technology in the control condition (random model).

Variables in the final model	Slope (b_Y)	SE(adj)	Lower 95th(adj)	Upper 95th(adj)	z-Value(adj)	p-Value	$Q_{Regression}$
Intercept	−0.23	0.29	−0.80	0.33	−0.81	.42	
Demographics – subject matter (reference category – non-STEM)							
D-1 STEM	−0.16	0.05	−0.26	−0.05	−2.88	.004	7.06*
Technology A – relative difference between the treatment and the control (reference category – high)							
D1-Low	0.24	0.13	−0.02	0.50	1.80	.07	
D2-Medium	0.14	0.14	−0.14	0.42	0.98	.33	11.71**
Technology B – pedagogical uses of technology (reference category – all other combos w/o CS)							
D-1 Communication support	0.40	0.31	−0.21	0.98	1.27	.21	
D-2 Search & retrieval support	0.60	0.34	−0.07	1.27	1.74	.08	
D-3 Presentation support (PS)	0.22	0.27	−0.31	0.74	0.81	.42	
D-4 Cognitive support (CS)	0.42	0.26	−0.10	0.93	1.59	.11	
D-5 Combinations (CS + PS)	0.45	0.26	−0.06	0.96	1.73	.08	
D-6 Combinations with CS	0.25	0.31	−0.35	0.85	0.81	.42	22.47***
$Q_{Residual}$							407.12****

*First step, df = 1, $p < .007$; **second Step, df = 3, $p = .008$; ***last step, df = 9, $p = .008$; ****residual, df = 390, $p = .27$.

Table 10
Levels of subject matter (mixed effects model).

Levels	Effect sizes			Standard error and 95th CI			Q-value
	g+(pred)	g+(unadj)	k	SE	Lower 95th	Upper 95th	
STEM	0.23	0.23	184	0.04	0.15	0.30	
Non-STEM	0.37	0.37	216	0.04	0.30	0.45	
Between							7.23*

*df = 1, $p = .007$.

Table 10 shows the comparison of STEM and non-STEM in some technology sub-collection. Table 11 details the three levels of difference between the treatment and the control. Post hoc analysis indicated that only low and high were significantly different ($z = 2.102$, $p < .05$).

Table 12 details the results of pedagogical uses of technology across all of its levels. As with the previous collection cognitive support outperformed presentation purpose ($z = 3.680$, $p < .001$), but in addition, combinations of CS and PS, based on a fairly large number of effect sizes, is equal to CS alone. This is a change from the previous collection where CS + PS was $g+ = 0.20$ based on 293 effect sizes. In fact, CS + PS is not significantly different from CS ($z = 0.065$, $p = .95$) and it is significantly different from PS ($z = 2.87$, $p = .004$). Interestingly, search and retrieval retain its highest position, but the mean is based on only five effect sizes and it is not significantly different from zero. However, its strong showing in both collections suggests an interesting avenue for future research. Clearly, there is something about this level that contributes to higher achievement.

3.3.8. *Variability in CS and PS distributions of effect sizes*

In examining the cognitive support (CS) and presentational support (PS) categories of primary purpose of technology use, it became apparent that, although CS produced significantly higher average effect sizes than PS in both sub-collections (i.e., no technology $g+ = 0.408$ vs. 0.167; and some technology $g+ = 0.335$ vs. 0.131), CS effect sizes were more variable than PS effect sizes. Table 13 shows three ways of judging this – the minimum and maximum effect sizes, the ranges of the effect sizes and their standard deviations. All of these indices point in the same direction, that cognitive support tools have higher highs and equal or lower lows than presentation support tools, in terms of their relationship to achievement outcomes. It is hard to envision a straightforward reason for these differences, but it may be possible that an answer lies with the ease or difficulty of successfully implementing tools in these categories. The possibility of this explanation will be explored in the Discussion.

Table 11
Levels of relative difference between the treatment and the control (mixed effects model).

Levels	Effect sizes			Standard error and 95th CI			Q-value
	g+(pred)	g+(unadj)	k	SE	Lower 95th	Upper 95th	
Low	0.34	0.34	306	0.03	0.30	0.40	
Medium	0.21	0.22	78	0.07	0.08	0.35	
High	0.13	0.14	16	0.14	−0.13	0.40	
Between							4.27*

*df = 2, $p < .119$.

Table 12
Levels of pedagogical uses of technology (mixed effects model).

Levels	Effect sizes			Standard error and 95th CI			Q-value
	$g+$(pred)	$g+$(unadj)	k	SE	Lower 95th	Upper 95th	
Communication	0.315	0.279	10	0.112	0.059	0.499	
Search & retrieval	0.569	0.571	5	0.382	−0.178	1.320	
Presentation (PS)	0.123	0.131	48	0.063	0.008	0.254	
Cognitive support (CS)	0.335	0.334	130	0.046	0.244	0.424	
Combination (CS + PS)	0.339	0.341	192	0.041	0.261	0.421	
Combinations with CS	0.143	0.144	11	0.222	−0.291	0.579	
Combinations w/o CS	−0.020	−0.096	4	0.093	−0.278	0.086	
Between							26.19*

*df = 6; $p = .0002$.

Table 13
Variability in cognitive support tools versus presentation support tools.

Pedagogical purposes	No technology in control			Some technology in control		
	Min/Max	Range	SD	Min/Max	Range	SD
Presentation support	−0.81/1.27	2.08	0.81	−1.19/1.24	2.43	0.85
Cognitive support	−0.85/2.30	3.15	1.23	−1.23/2.02	3.25	0.98

3.3.9. *Have the effects of technology implementation changed over time?*

Of interest in the final analysis was a question of whether the effects of technology on achievement had changed during the period covered by this meta-analysis. To do this, weighted simple regression was conducted, treating publication year as the predictor and g as the outcome variable. This was done for each sub-collection separately.

The results of the first analysis (i.e., No Technology in the Control condition) revealed that effects associated with technology use in postsecondary education ($k = 479$) have not changed substantially over the years covered by this meta-analysis ($b_Y = 0.004$, $p = .27$, $Q_{Regression} = 1.22$). The second analysis, of studies with Some technology in the Control ($k = 400$) yielded a similar result, that the effects of technology have not changed over time ($b_Y = 0.001$, $p = .91$, $Q_{Regression} = 0.01$).

3.4. *Attitude outcomes*

3.4.1. *Overall analysis*

Originally, 293 attitude effect sizes were extracted, 109 of which were based on pre-experimental designs. These low-quality studies were removed from the collection along with three extreme effects (i.e., $g > +2.5$), leaving 181 effect sizes for analysis.

In general, the magnitude of the weighted average effect size (Table 14) was somewhat lower for attitudes than for achievement outcomes ($g+ = 0.21$ vs. $g+ = 0.27$, respectively), but just as variable ($I^2 = 72.42$ for achievement vs. $I^2 = 77.45$ for attitudes). Under the fixed model, the final collection was judged to be significantly heterogeneous, with a moderate to high degree of between-study variation.

3.4.2. *Separate analyses*

The collection was divided into two sub-collections, as with achievement outcomes, studies with no technology and studies with some technology in the control condition. Tables 15 and 16 present the analytical details of these two sub-collections. It is notable that the weighted average effect size of the Some Technology in the Control Condition sub-collection (Table 16) is significant, but considerably lower than all of the other averages.

There were no significant moderator variables for either of the two sub-collections of studies.

4. Discussion

The overall question addressed in this review concerned the impact of technology on achievement and attitude outcomes in post-secondary classroom instruction. As noted in the Introduction, the two very large collections of effect sizes are somewhat different forms of the experimental research question "Does technology affect student achievement and attitudes"? The first type that we refer to in shorthand as "no technology" are studies in which the levels of the independent variable are some technology in the treatment condition and no

Table 14
Overall weighted average effect size (random effects model) for the final collection on attitudes and heterogeneity statistics (fixed effect model).

Population estimates	k	$g+$	SE	Lower 95th	Upper 95th
Final collection	181	0.21*	0.04	0.13	0.28
Heterogeneity analysis (fixed effect model)	$Q_{Total} = 798.23$ (df = 180), $p < .000$, $I^2 = 77.45$				

*$z = 5.25$, $p < .01$.

Table 15
Overall weighted average effect size (random effects model) for studies with no technology in the control group on attitudes and heterogeneity statistics (fixed effect model).

Population estimates	k	$g+$	SE	Lower 95th	Upper 95th
Sub-collection no technology in the control	102	0.27*	0.05	0.17	0.38
Heterogeneity analysis	$Q_{Total} = 486.256$ (df = 101), $p < .001$, $I^2 = 79.23$, $\tau^2 = 0.20$.				

*$z = 5.329$, $p < .001$.

technology in the control condition. This is the older design that has been reviewed in various meta-analyses and summarized by Tamim et al. (2011). Analysis of this sub-collection is intended to address questions Tamim's review could not address. The second type of design attempts to answer more nuanced questions about the efficacy of some technology in the control condition and some technology-based addition or modification in the treatment condition. Since the question asked and answered by these sub-collections, as well as the methodology for establishing the treatment and the control relationship, was somewhat different, we decided to analyze them separately. Each of the sections that follow will address the overall message that findings of this review convey along with the consistency or lack of consistency in these two sub-collections.

4.1. Overall findings

Overall, the sub-collections for both achievement and attitude outcomes are remarkably similar. Both sub-collections of achievement outcomes produced low to medium-low random model weighted average effect sizes ($g+ > 0.20$ and $g+ < 0.50$) that closely approximate the second-order findings of Tamim et al. (2011). Both sets were significant and both were heterogeneous under the fixed effect model. The outcomes for attitudes were generally the same but the some technology sub-collection produced an average weighted effect size that was less than half that of the no technology set of studies. This difference may suggest that using technology has greater effects than not using technology or using technology to varying degrees. It may also suggest that the finding in the no technology sub-collection is partly due to novelty or Hawthorne effects, as described by others (e.g., Kozma, 2001; Kulik, Schwalb, & Kulik, 1982).

4.2. Moderator variables in the achievement outcomes

4.2.1. Is Clark's claim still valid?

We will begin by examining Richard Clark's (1983, 1994) claim about the nature of the relationship between technology and pedagogy (i.e. instructional design). Clark argued that it is the nature of pedagogical strategies and designs that matter in the teaching/learning process and that:

> The best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition. Basically, the choice of vehicle might influence the cost or extent of distributing instruction, but only the content of the vehicle can influence achievement (1983, p. 445).

Our results partially support this position, but it is critical to understand that the crux of Richard Clark's (1983) original argument about the lack of impact of technology used in teaching was formulated during an era when technology was little more than presentational tools or tools intended to teach content directly (e.g., CAI) in postsecondary classrooms. Concordant with Clark, we found that technological applications that simply present information in an alternative form, such as PowerPoint, yield small effects, in our case a range from $g+ >0.10$ to $g+\ 0.20$. But even with these small effects, the average weighted effects in both sub-collections for presentation tools are significant, indicating that on average students in the treatment condition outperformed students in the control condition, albeit to a small degree. Technology did play some role in students' acquiring content knowledge, compared to the alternative in the control.

In the introduction of this piece the question was posed whether technology could affect measurable learning achievement in ways that do not involve imparting content. The answer to this question is yes, according to our findings. So, while Clark was essentially correct for a particular manifestation of technology use, available in the 1980s as well as today, his argument is less pertinent to the range of computer-based tools that are intended to broaden the reach and extend the learning capacity of newer generations of postsecondary students as both Cobb and Mayer explain. Our evidence shows that when more contemporary applications such as cognitive support are involved, more substantial effects emerged, within the $g+ >0.30$ to $g+ <0.45$ range.

Further evidence also supports this position. Although scant, based on only a few cases ($k = 4$ and 5, respectively), search and retrieval tools yielded large effect sizes ($g+ >0.50$ to $g+ <0.75$). To a lesser extent, tools used to help students communicate among themselves, with teachers and to interact with online content yielded similar outcomes ($k = 17$ and 10, respectively) with weighted average sizes in the between $g+ >0.20$ to $g+ <0.30$ range. These results are basically in line with findings by Bernard et al. (2009) who found similar patterns of results for interaction treatments in the distance education and online learning literature.

Table 16
Overall weighted average effect size (random effects model) for studies with some technology in the control group on attitudes and heterogeneity statistics (fixed effect model).

Population estimates	k	$g+$	SE	Lower 95th	Upper 95th
Sub-collection some technology in the control	79	0.11*	0.06	0.003	0.22
Heterogeneity analysis	$Q_{Total} = 298.99$ (df = 78), $p < .001$, $I^2 = 73.91$, $\tau^2 = 0.17$.				

*$z = 2.02$, $p = .043$.

Taken together, one fundamental point that Clark makes repeatedly will likely hold true, that tools, computer-based or otherwise, serve at the pleasure of the goals of learning (whether set by teachers or students themselves) and the design of the instruction into which they are intended to be integrated. In other words, it's the pedagogy not the technology that matters, although more correctly it is the synergistic relationship of these two. Technology makes possible strategies that could not otherwise be implemented. This relationship is discussed below in what the data suggest as the underlying factor, that being how technology supports passive to active learning as argued by Cobb and Mayer, among others.

4.2.2. Upper limits to technology use

A separate but related question asked and answered in moderator variable analysis was what is the difference in technology use between the treatment and the control conditions. Essentially, this is a question that is more absolute in the case of the no technology sub-collection (i.e., difference between *no technology* and *some technology*) and more relative in the case of some technology (i.e., difference between *some technology* and *some technology plus something else*). This means that the answers should be different and to some extent they are. For both sub-collections a high degree of difference resulted in lower weighted average effect sizes, in the range from $g+ > 0.05$ to $g+ < 0.13$. Thus, the sub-collections unanimously suggest that too much technology in the classroom could impair learning outcomes.

While the degree of difference in technology use in control and treatment is interesting and a potentially useful finding, especially since it favors low to moderate use of technology, the results of the variable pedagogical uses of technology may turn out to be the most intriguing.

4.2.3. The pros and cons of instructional technology variables

In discussing the findings described above (Section 4.2), one must always be cognizant of the large degree of variability surrounding the weighted average effect sizes. This variability says that some implementations of technology are better than others and that some actually affect student achievement deleteriously (i.e., negative effect sizes). The challenge of educational technology research, then, is uncovering the links and circumstances under which pedagogy and technology operate as effective partners, while avoiding the pitfalls that tend to reduce, negate or actually reverse effectiveness.

The results that underlie this discussion of the positives and negatives associated with various forms of pedagogical uses of technology is predicated on the differential degree of variability surrounding each average effect size as shown in Table 13 in the Results. This table shows much wider variability surrounding the category of cognitive support than presentation support across both sub-collections.

Learning with cognitive support tools involving strategies that include student-centered, dynamic, interactive techniques appears to produce larger effect sizes. Well-designed and implemented simulations, games, wikis, etc. encourage the types of active, problem/exploratory-based learning that decades of research have suggested as supporting higher levels of meaningful learning (Schwartz & Schmid, 2012). Presentation software, on the other hand, shows a much lower positive impact, but also a narrower distribution. These strategies are largely instructor-controlled, and often involve the passive dissemination and acquisition of information.

The following sections (Sections 4.2.4 & 4.2.5) discuss the circumstances under which cognitive tools and presentation software work, and fail to work. Because this was a meta-analysis, where by definition categories of moderators include a variety of instantiations of the category definition, one can best discuss these factors by using exemplars that illustrate the results and pedagogical implications emanating from these findings.

4.2.4. Cognitive support tools

Presentation and some communication techniques may serve to disseminate information, echoing Clark's argument, but cognitive tools focus on scaffolding the active creation and negotiation of information. These activities appear to make learning more efficient and more meaningful, consistent with arguments about technological efficiency (Cobb, 1997) and learning from multimedia (Mayer, 2008), especially interactive multimedia. However, their effects are also more variable and they can seemingly fail under certain conditions.

4.2.4.1. Positive examples. To illustrate the power of utilizing cognitive support, high positive effects were observed in many studies. The highest effect ($g = +2.30$) derived from the study by Issenberg et al. (2002) compared learning outcomes of 4th-year medical students either working with the UMedic multimedia software, a cardiology patient simulator (experimental group), or exposed to case-based in-class discussions (control group). Technology use in the experimental condition allowed students to study up to 27 realistically simulated common and rare cardiac conditions by observing the corresponding changes in symptomology, for example, … "blood pressure, bilateral jugular venous pulses, bilateral carotid and peripheral arterial pulses, precordial impulses in six different areas and auscultatory events …" (p. 225), making the learning experience more tangible (authentic). Instructor-led discussions in the control condition, though based on real patient cases, lacked the level of authentic interactivity provided by simulations in the experimental condition. The latter helped students to link easily observed and repeatable patterns in symptoms to particular cases they had studied.

As another example, Chen and Levinson (2006) incorporated a network growth simulator program (SONG) into a senior/graduate course involving transportation system analysis. The simulation effectively ($g = +0.73$) "enhanced students' learning in terms of helping students develop in-depth understanding about the development process of network patterns, and helped them develop some aspects of judgment, problem-solving, and decision-making skills" (p. 29). The simulation provided students with an interactive learning environment and access to diversified learning strategies, leading to higher levels of understanding. Chen and Levinson (2006) proposed a number of design implications for such approaches, including offering reasonable complexity and interactive and unambiguous feedback, guiding workload management, creating fun associated with manipulating dynamic variables, and establishing goals and access to supporting information to encourage discovery/exploration.

4.2.4.2. Negative examples. As a negative example, in a true experiment by Mäkitalo, Weinberger, Häkkinen, Järvelä, and Fischer (2004), students in an online collaborative environment received epistemic scripts, that is, computer-generated prompts that in the context of a particular problem reminded them of the related concepts they had learned previously. The treatment increased discourse, but decreased information-seeking activities, resulting in the unscripted condition (control) yielding superior learning outcomes ($g = -1.23$). It seems that learners in the control condition, who did not receive these prompts, made greater and more effective use of online discussion boards available to both groups. The experimental conditions relied on support provided by pre-constructed prompts, thus losing the dynamic quality of CMC and limiting the variety of feedback. Again, providing learners with increased flexibility and the opportunity to explore produces better learning gains. While technologically inviting and likely more accurate, pre-scripted feedback constrained active, meaningful learning.

Similarly, Darabi, Nelson, and Palanki (2007) reported much better achievement outcomes and mental effort for students in the control condition. They practiced troubleshooting malfunctions using a computer simulation of a water–alcohol distillation plant without supportive information of so-called worked examples of two types (process- and product-oriented) available to students in the experimental condition ($g = -1.10$). Though designed to remind learners of the most likely causes of the observed malfunctions or of the typical procedural steps for solving particular problems, worked examples impeded an examination of a broader variety of options in the simulation. This resulted in a lack of understanding the source of the problem, or finding the most relevant solution. The authors speculated that using worked examples without practice tasks interfered with learners' existing schema. They also state that the interactivity with authentic tasks implicated in all conditions was deemed a necessary ingredient in effective learning and transfer. In this case, it was the pedagogy rather than the technology that had a greater impact on learning. Conventional problem solving in a simulated environment, which was less directive than the worked examples, proved superior.

4.2.5. Presentation support tools

Presentation support tools yielded, lower overall average effect sizes than cognitive support tools, but the gap in magnitude between positive and negative average effects were not as broad as cognitive support. Interesting effects were observed on both poles of the distribution.

On the positive side, largely in line with Mayer's principles for the design of multimedia learning tools, Lewalter (2003) found that students who learned the topic of optical phenomena using computer-based text enhanced with visual illustrations (both dynamic and static) did much better on the posttest than the students of the test-only control group ($g = +1.14$). Interestingly, there was no difference between dynamic and static visuals.

On the negative side, Scheiter, Gerjets, Huk, Imhof, and Kammerer (2009), found large negative effect sizes, $g = -0.72$ and $g = -0.42$, across two studies when design issues were compared in high versus low technology presentation methodologies. University students watched either realistic or schematic visualizations of the process of mitosis as part of a biology course. The realistic condition consisted of videos of mitosis occurring under the microscope. The students in the low technology dynamic visualization condition viewed schematic drawings of the sequence of processes. While the technology provided a detailed, realistic representation of the phenomenon, the videos included irrelevant and dynamic information that distracted learners (e.g., irrelevant particles floating around the nucleus), and placed deleterious demands on their cognitive processing, even for high prior knowledge students.

Presentation applications appear to be more effective when extraneous information is stripped away and focus is centered on the essential aspects of the process being modeled. Too much information was not ideal; neither was too little. This finding is similar to what we observed with *degree of technology use*, i.e., too much technology resulted in significantly lower achievement ($g+ = +0.07$), whereas low and medium levels both proved to be equally effective ($g+ = +0.33$ and $+0.32$, respectively). Technology used for technology's sake often results in excessive information leading to cognitive overload which then can interfere with learning and performance (Dede, 2004).

4.2.6. Contextual moderator variables

There were two contextual moderator variables: subject matter and classroom/blended context. Only one of these, subject matter, was a significant predictor in multiple meta-regression in both sub-collections.

While the results of the analysis of STEM vs. non-STEM subject matters were significant, they produced contradictory findings across the two sub-collections. STEM subjects outperformed non-STEM subjects when there was no technology. The reverse occurred for the sub-collection of studies with some technology. Any attempt to explain this outcome would be speculative. The outcomes are nonetheless interesting, given the world-wide emphasis on STEM subjects (Gonzalez & Kuenzi, 2012). The some technology reversal calls for further research to examine the nature of precisely what pedagogical approaches are being employed, and why they appear to be less effective.

Overall, we did not find the variable classroom/blended context to be a significant predictor in either regression model, but upon examination of the level of that variable called *blended in the treatment/classroom in the control*, the results for achievement outcomes were significant (no technology, $g+ = 0.35$, $k = 100$, $p < .001$; some technology, $g+ = 0.28$, $k = 14$, $p < .01$). The first of these is similar to the meta-analysis conducted by Means, Toyama, Murphy, Bakia, and Jones (2010), who found a $g+$ of 0.35 between blended and non-blended instructional classroom conditions for achievement outcomes. We also found a similar result with attitude outcomes in the no technology studies ($g+ = 0.33$, $k = 21$, $p < .001$). There weren't enough effect sizes in the some technology categories for meaningful interpretation. All in all, these data suggest that blended approaches are more effective and favored. Similar to the STEM issue, this emerging pedagogical approach requires careful research vis-à-vis how the technology best serves learning and attitudes.

4.3. Conclusions

The overall message emerging from this study is that learning is best supported when the student is engaged in active, meaningful exercises via technological tools that provide cognitive support. But we are a long way from understanding more specifically how to design effective cognitive support tools and when precisely and how to integrate them into instruction. While Cobb (1997), Mayer (2008) and others focus on the information-processing aspects of cognitive tools, the majority of the studies reviewed also cited issues related to the motivational dispositions of learners, including the nature of students' learning goals, their expectations for success, and their causal attributions for learning outcomes in connection with the use of cognitive support tools (e.g., Abrami, Bernard, Bures, Borokhovski, & Tamim, 2011; Azevedo & Hadwin, 2005). The same sets of questions apply to presentation strategies, communications, and retrieval strategies. For example, the rapid introduction of tablets and smart phones, and expanded web-based conferencing will certainly change the landscape of educational interaction.

Decades of research on cognition and learning support the transformation of postsecondary classroom instruction away from purely didactic methods of conveying instructional content to more engaging and interactive methods that focus on maximizing students' learning.

Technology is demonstrably playing an important role in improving pedagogy. Our research and development agenda should now shift toward a more fine-grained analysis of identified instructional factors to better aid designers, software developers, teachers, researchers and theorists.

Acknowledgments

This research was supported by grants to Schmid, Bernard and Abrami from the Social Sciences and Humanities Research Council of Canada, and Fonds de recherche sur la société et la culture, Province of Québec. We would also like to acknowledge the work of David Pickup and Anna Sokolovskaya.

Appendix A

Codebook

Categories and details of the codebook
Study ID number
Sequential number of the ES
Author(s)

Publication Data
Year of publication
Type of publication

1) Journal (later to be further categorized by the impact factor)
2) Dissertation
3) Conference Proceedings
4) Report/Gray literature

Effect Size Extraction
***n* for the experimental group**
***n* for the control group**
Total *N* (the entire sample size)
Effect size (*d*)
Procedure of *ES* extraction

1) Calculated using reported descriptive statistics
2) Calculated using reported inferential statistics
3) Estimated from partial inferential statistics, e.g. reported p-value
4) Estimated from hypothesis ($p < \alpha$) or assumption of equal sample size when only N is given.
5) ES reported by the authors (only used when no other information is available)

Outcome Information
Outcome type

1) Achievement
2) Attitudes

Form of outcome measure

1) Most representative (cumulative) one-time performance measure
2) Reported composite of several tests/evaluations
3) Calculated composite of several assessments reported in the paper
4) Individual measure/item selected to represent the corresponding outcome

Nature of comparison – Brief description of both, experimental and control, conditions and of the source of data for ES extraction (open-ended entry)

Methodological Quality
Research design

1) Pre-experimental design (PED, e.g., one-group pretest/posttest or two-group w/o control for selection bias)[a]
2) Quasi-experimental design (QED, non-equivalent groups with control for selection bias, etc.)
3) True experimental design (RCT, random assignment of participants to groups, etc.)

Instructor equivalence

1) Same instructor for experimental & control groups
2) Different instructors for experimental & control groups

999 Unknown (missing)
Material equivalence (e.g., textbook, assignments, subject matter content equivalence)

1) Same set of materials for experimental & control groups
2) Different sets of materials for experimental & control groups

999 Unknown (missing)

Demographics
Academic level of learners

1) Undergraduates
2) Graduate students
3) Special programs (military, etc.)
4) Mixed groups

Subject matter[b]

1) Math (including stats and algebra)
2) Language (including language arts and second language learning)
3) Physical and natural science (including biology, physics, chemistry, geology)
4) Social sciences (including history, sociology, geography)
5) Psychology
6) Philosophy
7) Computer science (information technology)
8) Education
9) Health sciences (including medicine, environmental health, and nursing)
10) Business (including economics and management)
11) Engineering
12) Others (specify)

999 Unknown (Missing)

Nature of Treatment
Length of the treatment

1) One-shot (within a day) treatment (experiment)
2) Longer than one-shot (within a day), but less than a term (semester)
3) One term (One semester)
4) More than one semester (The duration of the treatment was longer than one term)

999 Unknown (missing)

Strength of the treatment – Magnitude of the difference between Experimental and Control conditions in the degree of technology use.

1) $G_E > G_C$
2) $G_E \gg G_C$
3) $G_E \ggg G_C$

Difference in degree of technology use

This item is intended to reflect the overall difference in the degree of technology use – to make the decision of selecting EXPERIMENTAL (as oppose to CONTROL) condition more transparent.

The codebook specifies that the difference between conditions in amount of technology may derive from one of the three major sources:

1) Intensity of technology use (EXPERIMENTAL group used technology more frequently or/and for longer time period). Comparisons between conditions using technology and those with no technology at all belong here.
2) Advancement (technology in EXPERIMENTAL condition was more advanced in a sense that it featured more options, enabled more different functions).
3) Degree of use (EXPERIMENTAL group was exposed to more different types of technology, so that we could talk about some summative effect of multiple technological tools).

Clarifying Dimension of Pedagogical Uses Technology:

<u>Communication Support:</u>

When students in one group have more opportunities to interact among themselves (or/and with their teacher) than students in the other group using various means of electronic communication and do it faster, more frequently or with several different technologies/communication channels (e.g., discussion boards in addition to e-mail), this group will be coded as higher on dimension of Immediacy of Communication.

<u>Cognitive Support:</u>

When one group is given more support for learning via technological tools that are designed and used to analyze, reorganize and restructure learning materials, synthesize information, manipulate parameters, clarify and linked concepts to be learned, etc. than the other group, it will receive higher rating on the dimension of Cognitive Support. This dimension depicts how flexible students are with given learning materials and typically implies high level of cognitive interaction with the content.

Whenever technology is designed and used to reduce extraneous (i.e., unnecessary, unrelated to major learning objectives) cognitive load and to boost germane cognitive load by refocusing attention on most relevant to learning objectives aspects of instruction (e.g., Sweller, 1999; Sweller, van Merriënboer, & Paas, 1998), it will be considered to represent the function of cognitive support. Cognitive support may be provided by various means: preventing working memory overload by properly combining channels (modalities) of presenting information in the design and implementation of educational multimedia (e.g., Mayer & Moreno, 2003), optimizing information to meet the limits of working memory capacity and facilitate structured transfer of information to long-term memory (e.g., Cooper, 1998), providing additional tools that facilitate completion of the tasks necessary but not central for the main instructional goals and many others – essentially, everything that prevents taking mental resources away from the primary learning objectives, everything that helps structure information around learning goals – knowledge to be acquired, problem to be solved or skills to be developed, everything that directs learning activities toward what is in the focus of the corresponding educational experience.

Examples of cognitive support include, but are not limited to, simulations and virtual labs, software that scaffolds learning models process and provides elaborate (adaptive) feedback, creates concept maps and promotes knowledge building, reduces cognitive load of non-essential (secondary to learning objectives) tasks, etc.

<u>Search and Retrieval Support:</u>

This category features tools that provide capabilities for knowledge seeking and retrieval. Examples include: access to web-links, search engines, databases, or other electronic resources, etc.

<u>Presentation Support:</u>

Technology in this category offer tools designed and used for the purposes of presenting learning materials. Examples include: PowerPoint-type software, various illustrations, static and moving images, videodisks, etc., whenever they are the major means for delivery of the instructional content.

Major pedagogical uses of technology
Instruction to coders – Whenever possible put the code for one of five functionality options – **3a, 3b, 3c, 3d, or 3e** (i.e., try to make judgment which function contributed or was intended to contribute the most to the difference between conditions). If more than one apply equally, put **Multiple:** & specify – e.g., **3a, 3d**. If neither is applicable, put **999**.
Amount of Blended in Experimental condition (subjective estimation)

1) Fully in-class (F2F): On-line components limited, maximum, to assignments and posted lecture notes;
2) In terms of representation of on-line student activities/instruction components – the condition is between option 1 (above) close to fully in-class & option 3 (below);
3) High blended approach: Close to 50% of on-line student activities/instruction components.
4) Short duration experiments fully conducted in (simulated) on-line conditions without any substantial role played by instructor;

999 No way to define the among the levels of blended in experimental condition

Amount of Blended in Control condition (subjective estimation)

1) Fully in-class (F2F): On-line components limited, maximum, to assignments and posted lecture notes. Strong reasons to believe that only F2F components were present;
2) There is some evidence that the condition presented a combination of online and F2F instructional components;
3) High blended approach: Close to 50% of on-line student activities/instruction components. Strong reasons to believe that the condition presented a significant integration of online instruction components;
4) Short duration experiments fully conducted in (simulated) on-line conditions without any substantial role played by instructor;

999 No way to define the among of blended in control condition.

[a] Studies in this category were eventually excluded from the analyses.
[b] Some individual subject matters were eventually combined to create larger categories for subsequent analyses (e.g., STEM).

Appendix B

Table B1
Percentage of different sources of studies.

Source of studies	Number of studies	Percentage
Journal articles	525	77.89
Dissertations	68	10.09
Conference proceedings	27	4.01
Reports, book chapters and other gray literature	54	8.01
Total	674	100.00

Table B2
Sources of included studies by journal with the corresponding frequencies.[a]

Journal name	Number of studies
Computers & Education (CAE)	28
Journal of Computer Assisted Learning (JCAL)	22
Journal of Educational Computing Research	21
British Journal of Educational Technology (BJET)	12
Medical Education	12
International Journal of Instructional Media	10
Journal of Educational Multimedia and Hypermedia	10
Language Learning & Technology	9
Learning and Instruction	9
Educational Technology Research and Development (ETR&D)	8
Journal of Educational Technology Systems	8
Journal of Dental Education	8
Instructional Science	7
Journal of Nursing Education	7
Computers in Human Behavior	7
Academic Medicine	7
CALICO Journal	6
Computer Assisted Language Learning (CALL)	6
Journal of Educational Psychology (JEP)	5
Australasian Journal of Educational Technology	5
Journal of Engineering Education	5
American Journal of Pharmaceutical Education	5
Journal of Computers in Mathematics and Science Teaching	5

Table B2 (*continued*)

Journal name	Number of studies
European Journal of Dental Education	5
Teaching and Learning in Medicine	5

[a] Journals with a frequency of less than 5 are not included in the above list. The total number of journals – sources of reviewed and included studies is 246. The list includes (but is not limited to): Canadian Journal of Educational Communication, Journal of Research on Computing in Education, Computers in Nursing, Educational and Training Technology International, Electronic Journal of e-Learning, Journal of College Science Teaching, Journal of Computer-Based Instruction, Journal of Computing in Higher Education, Behavior Research Methods, Instruments, & Computers, Teaching of Psychology, Research Strategies, Journal of the Learning Sciences, Journal of Research in Science Teaching, Cognitive Science, Contemporary Educational Psychology, American Journal of Occupational Therapy, International Journal of Science Education, Journal of Information Technology Education, Advances in Health Sciences Education, Applied Cognitive Psychology, Journal of Applied Behavior Analysis, British Journal of Educational Psychology, Journal of Science Education and Technology, Interactive Learning Environments, Journal of Chemical Education, College and Research Libraries, Computer Applications in Engineering Education, Journal of Research on Computing in Education, Studies in Health Technology and Informatics and many other journals.

Appendix C

Table C1
Categories, numbers, and percentages of excluded studies.

Reasons for study exclusion	Number	Percentage
No comparison group (i.e., one group studies)	615	22.44
Multiple reasons	289	10.54
Not a study of technology	253	9.23
Insufficient or no statistical information	245	8.94
No difference between conditions	197	7.19
Qualitative methodology	181	6.60
No applicable outcome measure	180	6.57
Distance Education Study	173	6.31
Descriptive or opinion piece	164	5.98
Wrong unit of analysis	153	5.58
K-12 (not postsecondary)	128	4.67
Non-institutional course or workshop	121	4.41
Meta-analysis or other review article	42	1.53
Total	2741	100.00

References*

Abrami, P. C., & Bernard, R. M. (2013). Statistical control versus classification of study quality in meta-analysis. *Effective Education*. http://dx.doi.org/10.1080/19415532.2012.761889.

Abrami, P. C., Bernard, R. M., Bures, E., Borokhovski, & Tamim, R. (2011). Interaction in distance education and online learning: using evidence and theory to improve practice. *Journal of Computing in Higher Education (Special Issue: Improving Interaction in Distance Education Through Research), 23*(2–3), 82–103. http://dx.doi.org/10.1007/s12528-011-9043-X.

Azevedo, R., & Hadwin, A. F. (2005). Introduction to special issue: scaffolding self-regulated learning and metacognition: implications for the design of computer-based scaffolds. *Instructional Science, 33*, 367–379. http://dx.doi.org/10.1007/s11251-005-1272-9.

*Barnes, W. G. W. Constructing knowledge from an ill-structured domain: testing a multimedia hamlet. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA, April 1994 (ERIC Document Reproduction Service no. ED372743).

*Behnke, C., & Ghiselli, R. (2004). A comparison of educational delivery techniques in a foodservice training environment. *Journal of Teaching in Travel & Tourism, 4*(1), 41–56. http://dx.doi.org/10.1300/J172v04n01_03.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, A., Tamim, R., Surkes, M., et al. (2009). A meta-analysis of three interaction treatments in distance education. *Review of Educational Research, 79*(3), 1243–1289. http://dx.doi.org/10.3102/0034654309333844.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare to classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379–439. http://dx.doi.org/10.3102/00346543074003379.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis Version 2*. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. (2008). *Comprehensive meta-analysis* (Version 2.2.048) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111. http://dx.doi.org/10.1002/jrsm.12.

Bransford, J. D., Brown, A. L., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Carpenter, C. R., & Greenhill, L. P. (1955). *An investigation of closed-circuit television for teaching university courses* (Report 1). University Park, PA: The Pennsylvania State University.

Carpenter, C. R., & Greenhill, L. P. (1958). *An investigation of closed-circuit television for teaching university courses* (Report 2). University Park, PA: The Pennsylvania State University.

*Chen, W., & Levinson, D. M. (2006). Effectiveness of learning transportation network growth through simulation. *Journal of Professional Issues in Engineering Education & Practice, 132*, 29–41. http://dx.doi.org/10.1061/(ASCE)1052-3928(2006)132:1(29).

Clark, R. C., & Mayer, R. E. (2011). *E-Learning and the science of instruction: Proven guidelines for consumers* (3rd ed.). Hoboken, NJ: Wiley.

* Note: References marked with stars (*) are studies in the meta-analysis used as examples in text.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*, 445–449. http://dx.doi.org/10.3102/00346543053004445.
Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–29. http://dx.doi.org/10.1007/BF02299088.
Cobb, T. (1997). Cognitive efficiency: toward a revised theory of media. *Educational Technology Research and Development, 45*(4), 21–35. http://dx.doi.org/10.1007/BF02299681.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
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. http://dx.doi.org/10.1080/01421590802691393.
Cooper, G. (1998). *Research into cognitive load theory and instructional design at UNSW*. Sydney, Australia: University of New South Wales (UNSW). Retrieved from http://dwb4.unl.edu/Diss/Cooper/UNSW.htm.
Cooper, H., & Koenka, A. C. (2012). The overview of reviews: unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist, 67*, 446–462. http://dx.doi.org/10.1037/a0027119.
*Darabi, A. A., Nelson, D. W., & Palanki, S. (2007). Acquisition of troubleshooting skills in a computer simulation: worked example vs. conventional problem solving instructional strategies. *Computers in Human Behavior, 23*, 1809–1819. http://dx.doi.org/10.1016/j.chb.2005.11.001.
Dede, C. (1996). Emerging technologies and distributed learning. *American Journal of Distance Education, 10*(2), 4–36. http://dx.doi.org/10.1.1.136.1029.
Dede, C. (2004). Enabling distributed learning communities via emerging technologies (Part One – September) *THE Journal, 32*(2), 12–22 http://eric.ed.gov/?id=EJ709366.
Fletcher-Flinn, C. M., & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): a meta-analysis. *Journal of Educational Computing Research, 12*, 219–242.
*Ford, B. A., & Klicka, M. A. (1994). *The effectiveness of computer assisted instruction supplemental to classroom instruction on achievement growth in courses of basic and intermediate algebra*. Newtown, PA: Bucks County Community College (ERIC Document Reproduction Service No. ED378044).
*Ford, N., & Chen, S. Y. (2000). Individual differences, hypermedia navigation, and learning: an empirical study. *Journal of Educational Multimedia and Hypermedia, 9*(4), 281–311.
Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
Gonzalez, H. B., & Kuenzi, J. J. (2012). *Science, technology, engineering and mathematics (STEM) education: A primer*. Washington: DC: Congressional Research Service. Retrieved on March 22, 2013 from http://www.fas.org/sgp/ers/misc/R42642.pdf.
Grabe, M., & Grabe, C. (2006). *Integrating technology for meaningful learning* (5th ed.). Boston, MA: Houghton Mifflin.
Haque, S., & Srinivasan, S. (2006). A meta-analysis of the training effectiveness of virtual reality surgical simulators. *IEEE Transactions on Information Technology in Biomedicine, 10*(1), 51–58. http://dx.doi.org/10.1109/TITB.2005.855529.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis*. Washington, DC: National Science Teachers Association.
Higgins, J. P. T., Thompson, S. G., Deeks, J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560. http://dx.doi.org/10.1136/bmj.327.7414.557.
Hsu, Y. C. (2003). *The effectiveness of computer-assisted instruction in statistics education: A meta-analysis* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3089963).
*Huang, J. C. y (1995). Digitized speech as feedback on cognitive aspects of psychomotor performance during computer-based instruction. In M. R. Simonson, & M. L. Anderson (Eds.), *Proceedings of the 1995 annual national convention of the association for educational communications and technology* (pp. 213–218). Washington, DC: AECT.
*Issenberg, S. B., McGaghie, W. C., Gordon, D. L., Symes, S., Petrusa, E. R., Hart, I. R., et al. (2002). Effectiveness of a cardiology review course for internal medicine residents using simulation technology and deliberate practice. *Teaching and Learning in Medicine: An International Journal, 14*(4), 223–228. http://dx.doi.org/10.1207/S15328015TLM1404_4.
*Jackson, J. W. (2002). Enhancing self-efficacy and learning performance. *Journal of Experimental Education, 70*(3), 243–254.
*Katayama, A. D., & Crooks, S. M. (2001). Examining the effects of notetaking format on achievement when students construct and study computerized notes. *Learning Assistance Review, 6*(1), 5–23.
Koufogiannakis, D., & Wiebe, N. (2006). Effective methods for teaching information literacy skills to undergraduate students: a systematic review and meta-analysis. *Evidence Based Library and Information Practice, 1*(3), 3–43. Retrieved from http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/76/153.
Kozma, R. (1991). Learning with media. *Review of Educational Research, 61*, 179–221. http://dx.doi.org/10.3102/00346543061002179.
Kozma, R. (1994). Will media influence learning: reframing the debate. *Educational Technology Research and Development, 42*(2), 7–19. http://dx.doi.org/10.1007/BF02299087.
Kozma, R. (2001). Learning with media. In D. Ely, & T. Plomp (Eds.), *Classic writings on instructional technology* (pp. 155–188). Colorado: Libraries Unlimited, Inc.
Kozma, R. (2003). Information and communication technologies and educational change: a global phenomena. In R. Kozma (Ed.), *Technology, innovation, and educational change: A global perspective*. Eugene, OR: International Society for Educational Technology.
Kulik, C. C., Schwalb, B. J., & Kulik, J. A. (1982). Programmed instruction in secondary education: a meta-analysis of evaluation findings. *Journal of Educational Research, 75*(3), 133–138.
*Lewalter, D. (2003). Cognitive strategies for learning from static and dynamic visuals. *Learning and Instruction, 13*, 177–189. http://dx.doi.org/10.1016/S0959-4752(02)00019-1.
Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: good, bad, and ugly. *The Annals of the American Academy of Political and Social Science, 587*, 69–81. http://dx.doi.org/10.1177/0002716202250791.
*Mäkitalo, K., Weinberger, A., Häkkinen, P., Järvelä, S., & Fischer, F. (2004). Epistemic cooperation scripts in online learning environments: fostering learning by reducing uncertainty in discourse? *Computers in Human Behavior, 21*, 603–622. http://dx.doi.org/10.1016/j.chb.2004.10.033.
Mayer, R. E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice-Hall.
Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52. http://dx.doi.org/10.1207/S15326985EP3801_6.
Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Technical Report. Washington, DC: U.S. Department of Education.
Michko, G. M. (2007). *A meta-analysis of the effects of teaching and learning with technology on student outcomes in undergraduate engineering education* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3089963).
Murray, O. T., & Olcese, N. R. (2011). Teaching and learning with iPads, ready or not? *TechTrends, 55*(6), 42–48. http://dx.doi.org/10.1007/s11528-011-0540-6.
O'Donnell, A. M., D'Amico, M., Schmid, R. F., Reeve, J., & Smith, J. K. (2008). *Educational psychology: Reflection for action* (Canadian ed.). Mississauga, ON: John Wiley and Sons Canada.
*Pemberton, J. R., Borrego, J. J., & Cohen, L. M. (2006). Using interactive computer technology to enhance learning. *Teaching of Psychology, 33*, 145–147. http://dx.doi.org/10.1207/s15328023top3302_9.
Pigott, T. (2012). *Advances in meta-analysis*. New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-2278-5.
Ross, S. M., Morrison, G. R., & Lowther, D. L. (2010). Educational technology research past and present: balancing rigor and relevance to impact school learning. *Contemporary Educational Technology, 1*, 17–35. Retrieved from http://www.cedtech.net/articles/112.pdf.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis prevention, assessment and adjustments*. Chichester, UK: John Wiley & Sons.
*Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction, 19*, 481–494. http://dx.doi.org/10.1016/j.learninstruc.2008.08.001.
Schenker, J. D. (2007). *The effectiveness of technology use in statistics instruction in higher education: A meta-analysis using hierarchical linear modeling* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3286857).
Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R., Abrami, P. C., Wade, C. A., et al. (2009). Technology's effect on achievement in higher education: a stage I meta-analysis of classroom applications. *Journal of Computing in Higher Education, 21*(2), 95–109. http://dx.doi.org/10.1007/s12528-009-9021-8.
Schwartz, N. H., & Schmid, R. F. (2012). Using technology to foster meaningful learning environment. In M. J. Lawson, & J. R. Kirby (Eds.), *The quality of learning: Dispositions, instruction, and mental structures*. Cambridge, UK: Cambridge University Press.
Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.
Sweller, J., van Merrienboer, J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. http://dx.doi.org/10.1023/A:1022193728205.
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*(3), 4–28. http://dx.doi.org/10.3102/0034654310393361.
Timmerman, C. E., & Kruepke, K. A. (2006). Computer-assisted instruction, media richness, and college student performance. *Communication Education, 55*, 73–104. http://dx.doi.org/10.1080/03634520500489666.

Valentine, J. C., & Cooper, H. (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. http://dx.doi.org/10.1037/1082-989X.13.2.130.

Weston, M. E., & Bain, A. (2010). The end of techno-critique: the naked truth about 1:1 laptop initiatives and educational change. *Journal of Technology, Learning, and Assessment, 9*(6). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1611/.

*Yu, F. Y., & Yu, H. J. J. (2002). Incorporating e-mail into the learning process: its impact on student academic achievement and attitudes. *Computers & Education, 38*, 117–126. http://dx.doi.org/10.1016/S0360-1315(01)00085-9.

Zhao, Y. (2003). Recent developments in technology and language learning: a literature review and meta-analysis. *CALICO Journal*, *21*(10), 7–27. Retrieved from https://www.calico.org/html/article_279.pdf.