COMMONWEALTH *of* LEARNING

Technology Application in Teaching and Learning:

SECOND-ORDER REVIEW OF META-ANALYSES

TECHNOLOGY APPLICATION IN TEACHING AND LEARNING:

Second-Order Review of Meta-analyses

Commonwealth of Learning

The Commonwealth of Learning (COL) is an intergovernmental organisation created by Commonwealth Heads of Government to promote the development and sharing of open learning and distance education knowledge, resources and technologies.

Technology Application in Teaching and Learning: Second-Order Review of Meta-analyses

Authors: Eugene Borokhovski, Rana Tamim and David Pickup

ISBN: 978-1-7772648-7-1

DOI: https://doi.org/10.56059/11599/4069

Published by:

COMMONWEALTH *of* LEARNING

4710 Kingsway, Suite 2500, Burnaby,
British Columbia Canada V5H 4M2
Telephone: +1 604 775 8200
Fax: +1 604 775 8210

Web: www.col.org
Email: info@col.org

Contents

Tables

Figures

List of Abbreviations

AI	artificial intelligence
ALEKS	Assessment and Learning in Knowledge Spaces
AR	augmented reality
ARS	audience response system
BL	blended learning
CAI	computer-assisted instruction
CALL	computer-assisted language learning
CI	classroom instruction
CMC	computer-mediated communication
COL	Commonwealth of Learning
CSCL	computer-supported collaborative learning
CT	computer technology
DE	distance education
DGBL	digital game-based learning
DGS	dynamic geometry software
ECE	exploratory computerised environments
EFL	English as a foreign language
ES	effect size
FCM	flipped classroom model
HMD	head-mounted displays
ICT	information and communications technology
ITS	intelligent tutoring systems
IVR	immersive virtual reality
LD	learning disability
MaI	meta-analysis
MALL	mobile assisted language learning
MCSCL	mobile computer-supported collaborative learning
MMRQG	meta-analysis methodological reporting quality guide
PDA	personal digital assistant
SCMC	synchronous computer-mediated communication
SLA	second language acquisition
STEM	science, technology, engineering and mathematics
VR	virtual reality
WBI	web-based instruction

Executive Summary

This second-order meta-analysis (i.e., systematic quantitative synthesis of individual meta-analyses) summarises how the use of technology affects learning (achievement outcomes) in three different educational settings: in-class, online and blended learning. Comprehensive literature searches identified 915 potentially relevant publications. Rigorous standard review procedures (including the removal of duplicates, independent double-coding and assessment of methodological quality) resulted in the admission of 131 primary meta-analyses (featuring 134 independent effect sizes) across settings, technology types, grade levels and subject matters. Aggregated effect sizes were organised in three independent collections around the setting type/delivery mode (in-class, online and blended), while technology type and major functionality, grade level and subject matter, plus publication date, coverage and representativeness, as well as some aspects of implementation quality of the included meta-analyses were all coded and analysed as moderator variables. Additional sensitivity analysis and analyses of methodological moderator variables further reduced the data set to 118 effect sizes distributed across educational settings' type as follows:

- In-class technology integration: 94 effect sizes
- Online learning: 11 effect sizes
- Blended learning: 13 effect sizes

A degree of publication bias was detected for the In-Class Technology Integration collection, and a minor adjustment to the overall weighted average effect size was subsequently implemented. No adjustments were needed for the other two collections, but they each contained one outlier. Both outliers were removed. Thus, the final effects for all three collections, reported below, represent the results both before and after these statistical corrections. Specifically, the overall weighted average effect size[1] for in-class technology integration was $g{++} = 0.415$, $p < .01$, $k = 94$, before the adjustment and $g{++} = 0.347$, $p < .01$, $k = 111$ (94 + 17), with the imputed ($k = 17$) effects, whereas online learning and blended learning meta-analyses produced effects of $g{++} = 0.169$, $p < .05$, $k = 11$ ($g{++} = 0.085$, ns, $k = 10$) and $g{++} = 0.470$, $k = 13$ ($g{++} = 0.385$, $p < .01$, $k = 12$), respectively (adjustments via the removal of outliers results are in parentheses). All these are reported according to the random effects analytical model (Riley et al., 2011).

Here are several noteworthy observations based on these findings and on the results of the subsequent moderator variable analyses:

- The methodological quality of meta-analyses is important. Meta-analyses that are less rigorous and less comprehensive tend to produce less reliable (inflated) effect sizes.

[1] In this publication, the designation $g{++}$ is used to depict the overall weighted average effect size of a second-order meta-analysis to separate it from $g{+}$ which is used to present aggregated results of a first-order meta-analysis and from Hedges' g calculated for individual primary studies.

- On average, in-class technology integration has a small to moderate effect on learning (the distribution of effect sizes is significantly heterogeneous — i.e., there are more and less effective technology-based instructional interventions).
- Compared to in-class technology integration, online learning is only marginally effective. Furthermore, it could be quite challenging to implement properly and is associated with potentially negative side effects (e.g., social isolation, risk of widening the digital divide, higher attrition rates).
- Blended learning is the most effective approach in terms of learning outcomes but is no less challenging than other approaches to implement properly (infrastructure, logistics, instructional design, etc.).
- Educational technology appears to be used not only more frequently for language (especially second language) learning than for other subjects, but also more successfully in terms of learning outcomes.
- Contrary to some previous findings, this second-order meta-analysis detected a slight change in effect sizes over time — namely, the trend for more recent meta-analyses to produce relatively higher achievement effect sizes.

Two relatively small but distinctive collections of meta-analyses were outside the scope of the current second-order meta-analysis and merit examination in the future:

1. those focusing on special needs populations (e.g., students with learning disabilities)
2. those that compare technology use in experimental conditions where additional factors are applied (e.g., interactive or self-regulated scaffolding, embedded pedagogical agents) with the unmodified use of the same technology in control conditions

The latter category of studies, which was identified in our searches and has yet to be explored in-depth and summarised, could be of particular interest to educational practitioners who want to move beyond the question "Does the technology work?" towards the more refined and targeted "How can we use educational technology more effectively?"

Introduction

The widespread effective implementation and use of information and communications technology (ICT) is crucial if countries across the world are to meet the demands and expectations of the 21st-century globalised economy and our increasingly digitised society — not only in the workplace but also in the promotion of political accountability and changes in educational opportunities for children and youth (Sanders & George, 2017). As Korunka and Hoonakker (2014) state, “There is no doubt that the development and implementation of information and communication technology during the last decade has had — and still has — a major impact on all levels of society” (p. 1). Not only have the media, manufacturing industries and commerce become increasingly technology-oriented (Turban et al., 2018) but public education too has begun to integrate ICT into educational settings of all kinds to the point where it is increasingly difficult to envision classrooms free of at least some form of modern digital technology. In particular, Internet-reliant technology is one of the most prominent means for the delivery of various forms of distance education and blended learning.

Most teachers in the 21st century do not question the need to integrate digital technologies into their teaching practices, but few realise that computers first appeared in educational contexts in the United States as far back as 1963, thanks to the *Vocational Education Act of 1963* (US Department of Health, Education & Welfare, 1965). Around the same time, the notion of individualised instruction with the concept of computer-assisted instruction (CAI) was introduced. Since then, computer technology has progressed through different developmental stages, numerous calls to integrate it into educational contexts, various waves of promised transformational power (not all of which were completely fulfilled) and many sceptical assessments of its capacity to advance teaching and learning. However, as technological advancements and innovations continued to progress at an unprecedented pace at the end of the 20th century and into the 21st, technology gradually became an everyday tool for teaching and learning in one form or another. Needless to say, the COVID-19 pandemic, and the global transition to emergency remote learning, cemented the technology’s status as a go-to learning tool. The pandemic has increased reliance on digital tools for education, putting more pressure on educators to acquire a better understanding of the nuances of educational technology and how it can be used for effective and efficient teaching and learning.

Although there has been a gradual return to face-to-face learning delivery in different educational contexts around the globe, the benefits accrued from technology-supported learning — including learner engagement, increased learner-centredness and more flexible and responsive accessibility — are being perceived more positively by various stakeholders in the education field. While some academics claim that the pandemic may have ushered in the start of a new era in education (see, for example, Azorín, 2020), others suggest that it will not lead to a totally new educational system but that it will “leave a lasting trace” (Daniel, 2020, p. 95). Daniel observes that the use of online learning will continue to expand in higher education and that K–12 education will invest further in various aspects of technology-based learning. As such, learning from previous research on best practices for technology use through rigorous syntheses will be crucial as policymakers plan for the future of academic institutions.

Below, we will briefly consider what is currently known about the effectiveness of computer-based technology applications in all three broad types of educational setting: classroom learning, online learning and blended learning.

In-Class Technology Integration

The impact and reach of technology are both significant and wide-ranging. Many people rely to at least some extent on computers for various aspects of their daily personal and professional life, but that does not mean that the topic is not controversial in an educational context.

One of the most influential early debates about educational technology, for example, goes back to Clark (1983), who insisted that it should play a rather auxiliary role (i.e., it does affect the quality of the education it delivers, and more conventional means of instruction would be equally effective), while others (e.g., Kozma, 1991) argued that its role in education is more substantive and transformative. The disagreement is probably rooted in the history of educational technology itself. Originally, technology was used almost exclusively to deliver instructional content, and as a medium it was no more effective than a human teacher, even a not especially experienced one. For example, early studies on distributed closed-circuit television teaching versus live teaching (Carpenter & Greenhill, 1995) found no differences between live teachers and televised teachers. Even the development of much more sophisticated computer tools and applications (e.g., computer-assisted learning, multimedia, hypermedia) did not lead to sufficiently improved student learning outcomes for the issue to be considered and resolved unequivocally in favour of educational technology.

However, the arrival of what Jonassen (1995) referred to as computer-based cognitive tools appeared to trigger a change in focus from what can be achieved by using multiple alternative media (i.e., Clark's argument) to what can be achieved primarily through them (i.e., Kozma's argument). Computer-based communication, simulations, serious games, blogs and wikis, social networking, search and retrieval platforms and applications, etc. promise unique benefits that go well beyond the simple transfer of content from a teacher to students. Add to this productivity software like spreadsheets, statistical packages, concept mapping programs and a host of other student-oriented applications and we can see that Clark's arguments, while still valid in certain domains, must be considered insufficient for examining both the overall benefits and potential deficits of the introduction and the subsequent use of computing in education.

Before researchers turned their attention to the issues embedded in the debate between Clark and Kozma and examined them through more refined research questions, it might have been expected that they would draw on several decades of primary research to try to determine the overall effect of educational technology on student learning achievements in comparison with instructional practices that did not use any technological tools. By the first decade of the 21st century the number of both primary empirical studies and various types of systematic review (including many meta-analyses) in the field of educational technology had grown significantly, but all focused on a nearly identical research question: How effective are classrooms that use technology compared to classrooms that are technology-free? Navigating and meaningfully interpreting this vast collection of research was a real challenge, primarily because of the sheer volume of information, and required an innovative and yet reliable approach to summarise important findings. Tamim et al. (2011) undertook a second-order meta-analysis (i.e., a meta-analysis of previous meta-analyses) to review 40 years of research that summarised the impact of computer-based technology on learning.

The second-order meta-analysis approach offers the potential of aggregating data from a growing body of meta-analyses harvested over several years, in the same way that meta-analyses attempt to reach more reliable and generalisable inferences than individual primary studies (see, for example, Peterson, 2001). In this second-order meta-analysis, 25 previously published meta-analyses that cut across all levels of formal education, subject areas and technology types, from the 1970s onward, were selected from a pool of about 75 studies (using the criterion of the lowest possible overlap among primary empirical studies) and their results were synthesised. The random effects model analysis revealed a weighted average effect size of $g{++} = 0.35$ ($p < 0.01$), encompassing 1,055 primary studies and 109,700 participants. Moreover, this result was further validated through identifying and reanalysing independent studies from the meta-analyses that reported references to their respective "includes" (there were 574 independent effect sizes with the total number of students being 60,853) in the standard meta-analytical procedure. The overall weighted average effect size of this validation study (also according to the random effects model) was nearly identical to the one revealed in the second-order analysis: $g{+} = 0.33$ ($p < 0.01$), which provided extra credibility. In addition, two moderator variable analyses (mixed model) produced statistically significant outcomes. Technology use in K–12 formal educational settings (i.e., from kindergarten to final school grade before entering post-secondary education for learners ranging in age from roughly 5–6 to 17–18 years) seemed to have a higher effect on learning ($g{+} = 0.40$, $k = 9$) than it did in post-secondary education settings ($g{+} = 0.29$, $k = 11$): $Q_{Between} = 4.83$ ($p < 0.05$), and technology use in a general category "for instructional support" was more effective ($g{+} = 0.42$, $k = 10$) than it was in the "for direct instruction" (i.e., content delivery) category ($g{+} = 0.31$, $k = 15$): $Q_{Between} = 3.86$ ($p < 0.05$). The latter result is of particular interest as it could further inform practice (i.e., what kinds of pedagogical use of technology and various potential blends thereof are more promising) and open the whole line of research on major purposes (i.e., key applied functions) of educational technology use (discussed in more detail later in this publication).

There was another important outcome which, although it did not reach the level of statistical significance, clearly demonstrated a linear tendency for meta-analyses of lower methodological quality (e.g., those that included studies of pre-experimental research design or were implemented with some noticeable deviation from best meta-analytical practices as outlined in, for example, Cooper, 2017, or Tamim et al., 2021) to produce larger effect sizes, whereas effects of more rigorously implemented methodology tended to be more modest in magnitude (see page 14). Thus, Tamim et al. (2011) addressed the question of how effective classrooms that use technology are compared to those that are technology-free through the use of meta-analyses that looked at the technology versus no technology question.

Tamim et al. (2011) concluded that, generally speaking, digital educational technology does enhance learning, even if only to a relatively small extent. However, the second-order meta-analysis could not be considered the final word on the topic of technology integration[2] in education for two major reasons: A second-order analysis, in general, as pointed out by Cooper and Koenka (2012), is limited in its ability to address the host of peripheral questions that can only be settled in a primary meta-analysis, where coding decisions are made by the meta-analyst based on detailed descriptions of various study characteristics supposedly provided by the authors of primary research, and the synthesis is conducted at the more granular level of the individual effect sizes. As a result, Tamim et al. (2011) could not examine several substantive and demographic moderator variables and thus provided insufficient texture or nuance to the overall results.

Another challenge that could not be fully addressed by this second-order meta-analysis related to the rapidly changing educational landscape, as new technological tools and applications tend to be introduced to instructional practices in different amounts, in different combinations and under varying circumstances, meaning that the classical comparison type of "technology versus no technology" is no longer appropriate. Comparisons of "technology versus technology" (more versus less and/or different types, as reflected in, for example, Schmid et al., 2014, who reported 400 effects of that type of comparison separately from 479 effects for more traditional "technology versus no technology" type of comparisons) are now much more frequently in the empirical research and have become the basis of a more reasonable and more pressing question in modern education.

The literature on educational technology, especially questions regarding the effectiveness of various technologies used for educational purposes versus the non-use of technology to achieve similar instructional objectives, is replete with primary studies, and the number continues to grow. Not surprisingly, several meta-analyses have appeared in the literature to address these primary studies around different technologies used with different learners for teaching various content. In just one systematic review by Bernard et al. (2018), the researchers found 49 such meta-analyses (with no technology in the control condition and not accounting for either distance or hybrid models of education), dated from 1982 to 2014. They identified 20 top-quality meta-analyses using an instrument for meta-analysis quality assessment (MMRQG – meta-analysis methodological reporting quality guide).

[2] The term "technology integration" is used in this report to underline the context of in-class instruction as opposed to the more generic term "technology use" that applies across delivery modes.

Table 1 shows these 20 meta-analyses with some of their key descriptors, including publication date, grade level, overall effect size, etc.

Table 1. Twenty best-quality meta-analyses from Bernard et al. (2018)

Meta-analyses	Publication year	MMQRG average	Educational level	Effect size (ES) ($g+$)	SE of $g+$	k
Schmid et al.	2014	1.91	HE	0.25	0.05	479
Cheung & Slavin	2011	1.77	K-12	0.16	0.15	42
Cheung & Slavin	2012	1.64	K-12	0.15	0.16	37
Goldberg et al.	2004	1.45	HE	0.41	0.35	8
Sosa et al.	2011	1.36	HE	0.33	0.21	23
Bayraktar	2000	1.32	HE	0.27	0.14	54
Hsu	2003	1.32	HE	0.43	0.25	16
Kuchler	1999	1.32	HE	0.44	0.18	33
Torgerson & Elbourne	2002	1.32	K–12	0.37	0.37	3
Schenker	2007	1.27	HE	0.24	0.24	18
Onuoha	2007	1.27	HE	0.26	0.24	57
Michko	2007	1.23	HE	0.43	0.13	62
Sitzman et al.	2011	1.18	HE	0.28	0.22	20
Yaakub	1998	1.18	HE	0.35	0.35	14
Grgurovic et al.	2013	1.14	HE	0.24	0.19	26
Timmerman & Kruepke	2006	1.14	HE	0.24	0.13	59
Van Lejeune	2002	1.09	HE	0.38	0.31	33
Tekbiyik & Akdeniz	2010	1.09	All grades	1.12	0.20	33
Soe et al.	2000	1.00	K-12	0.26	0.33	9
Pearson	2005	1.00	K-12	0.49	0.15	45

HE: Higher education
k: Number of effect sizes reported in the studies

A careful review of these meta-analyses resulted in the following findings: The overall weighted average effect size was $g{+}{+} = 0.29$ ($k = 20$), slightly lower than the summary of 76 technology integration meta-analyses summarised in Hattie (2008), without accounting for their methodological quality. Bernard et al. (2018) concluded that "the rapid evolution of technological tools (and vivid attention to and high hopes for them) is not exactly matched by adequate instructional design efforts and educational practice" (p. 18). For that reason, and because since 2018 more educational technology meta-analyses have entered the research literature, a new comprehensive and non-selective summary was needed.

Online and Blended Learning

The use of distance education to deliver instruction to remote students has long been a viable alternative to standard classroom instruction. For many years the delivery of such instruction was done through the mail (often called asynchronous distance education because there was little real-time or simultaneous contact between student and teacher or among students); it was later done through videoconferencing (often referred to as synchronous), whereby two classrooms would be linked and participate together simultaneously (see Bernard, Abrami, Lou et al., 2004, for the most comprehensive review of both options from the early 21st century). Neither of these two options is prominent today outside of dedicated distance education institutions (e.g., Indira Gandhi National Open University, Athabasca University), as various Internet-based applications have since become popular. Online learning happens exclusively off-campus but can be either synchronous or asynchronous, and blended learning is a combination of online and in-class instruction (e.g., online learning supplemented with some classroom-based components). In the early 2000s the "flipped classroom" was introduced as a variant of blended learning (see Bishop & Verleger, 2013) with specific activities, conducted in a specified order and for a specified time, for both the online and classroom portions. The design and organisation of these online and classroom components distinguishes blended learning from flipped classroom learning. So far, flipped classrooms appear to be more prevalent in higher education — and primarily in science, technology, engineering and mathematics (STEM), medical science and other technical subjects (Hew et al., 2021) — whereas blended learning courses are found at all educational levels (i.e., elementary school through higher education) across a wider range of subjects.

Table 2 shows a nearly complete picture of meta-analyses done in the areas of online learning, blended learning and flipped classrooms since 2000 (note that many more meta-analyses of flipped learning than online learning and standard blended learning have been conducted, especially in the health sciences and university-level sciences). Some of these meta-analyses are summarised in a second-order meta-analysis that is the most recent work in this table (Hew et al., 2021) and whose content is organised in descending chronological order.

The results of three representative meta-analyses of online learning versus classroom instruction (as a control condition) from 2007 to 2013 show only a slight improvement in online learning compared to classroom instruction by itself. Cook et al.'s (2008) study was based on the experience of healthcare workers and produced a mean of $g+ = 0.12$ from 63 studies, whereas Jahng et al.'s (2007) and Means et al.'s (2013) studies were based on the experience of university students and registered average effect sizes of 0.05 and below. Essentially, the delivery of online learning did not change the effectiveness of instruction substantially compared to classroom/control, except in the domains of medical education.

Table 2. Meta-analyses of distance, online, blended and flipped classroom learning vs classroom instruction since 2004

Authors & publication year	Inclusive years	Content/learner population	Distance education context	Number of effect sizes/ studies (ESs)	Mean/ sig. (*$p \leq .05$)
[1]Hew et al. (2021)	2018–2020	Health, science & combo, higher education	Flipped classroom (blended learning)	15 meta-analyses	0.37*
[1]Hew et al. (2020)	2008–2017	STEM, higher education	Flipped classroom (blended learning)	10 meta-analyses	0.49*
Sparkes (2020)	2000–2017+	Combo, higher education	Flipped classroom (blended learning)	125+ ESs	0.30*
Vo et al. (2017)	2001+	Combo, higher education	Blended learning	51 ESs	0.39*
Liu et al. (2016)	Through 2014	Health professions, adults	Blended learning	56 ESs	0.81*
Spanjers et al. (2015)	N/S	All grades	Blended learning	24 studies	0.34*
Bernard, Borokhovski, Schmid, Tamim & Abrami (2014)	2000–2010	Higher education	Blended learning	117 ESs	0.33*
Means et al. (2013/2009)	1996–2008	Higher education	Blended learning	23 ESs	0.35*
			Online learning	27 ESs	0.05
Cook et al. (2008)	1990–2007	Health workers	Online learning	63 studies	0.12*
Jahng et al. (2007)	1995–2004	Higher education	Online learning	20 ESs	0.02
Sitzmann et al. (2006)	1996–2005	Adults	Online learning	71 ESs	0.15
Williams (2006)	1990–2003	Health workers	Online learning	34 ESs	0.15
Zhao et al. (2005)	1966–2002	Higher education	Online learning	98 ESs	0.10
Cavanaugh et al. (2004)	1999–2004	K–12	Online learning	116 ESs	0.03
Bernard, Abrami, Wade et al. (2004)	1985–2002	All learners	[2]Asynchronous DE	174 ESs	0.05*

NOTES:
1 Second-order meta-analyses of flipped classroom studies (Hew et al., 2020; Hew et al., 2021).
2 This is a distinction between synchronous and asynchronous DE.
DE: distance education
ES: effect size
N/S: not specified

The most comprehensive review (Bernard, Abrami, Lou et al., 2004) stated that, in terms of its effect on learning outcomes, distance education is no different from classroom instruction ($g+ = 0.01$, ns), though asynchronous and synchronous modes of distance education produced significantly different average effect sizes in favour of the former, while attrition and attitudinal outcomes were somewhat negative. This meta-analysis of distance education applications prior to 2000 is listed at the bottom of the table for the sake of comparison. The results in Table 2 suggest, however, that blended learning offers a substantial improvement over both classroom instruction and dedicated online learning. Five meta-analyses (Bernard, Borokhovski, Schmid, Tamim & Abrami, 2014; Liu et al., 2016; Means et al., 2013; Spanjers et al., 2015; Vo et al., 2017) found means that are in the 0.30–0.39 range. Liu et al. found a higher-than-average effect size (i.e., $d = 0.81$, $k = 56$, $p < .001$) for blended courses in the health professions. However, online learning was sometimes used as the comparison condition instead of classroom instruction, so this study may not compare robustly with other meta-analyses of blended learning.

For flipped classrooms, many meta-analyses have been conducted since 2016, but only three are shown in Table 2. Two are second-order meta-analyses (2020 and 2021), encompassing 25 studies in total. All the meta-analyses in these two compilations concern STEM or technical subjects in higher education. One first-order meta-analysis (Sparkes, 2020) that spans a greater range of subject matters in higher education is included. The results are a little lower, but still essentially the same as those from the more technical STEM or health sciences areas.

Based on the overall picture presented here, it seems that there is a fairly clear breakpoint between courses that are fully online and courses that include some component of classroom-based activity mixed with online learning. Clearly, there is something to be gained from combining some form of online with some form of classroom-based instruction. So far, we do not know what that *something* is, which creates a challenge in terms of investigating what composition of online and in-class instructional strategies/activities in what proportions and for what specific teaching/learning purposes works best.

All the evidence accumulated to date indicates consistently that online learning is a viable, though not necessarily superior, alternative to traditional classroom instruction, whereas blended and/or flipped learning appear to be a highly promising compromise between the two. However, this conclusion is rather generic, and the applications of either form of non-classroom educational practices in specific settings or for specific populations of learners deserve more detailed examination.

Research Questions

The main purpose of this review is to systematically identify meta-analyses that address the effectiveness — that is, the primary impact on learning achievement outcomes — of computing (digital) technology integration in education, encompassing all delivery modes and formats. These delivery modes and formats include classroom, distance, online and blended learning (combined classroom and online interventions) in formal educational settings. In addition, the objective is to summarise their respective findings in the form of a second-order meta-analysis (Tamim et al., 2011). The following major research questions guided this review:

1. Based on a systematic summary of previously published meta-analyses, what is the overall weighted average effect size (as the comprehensive point estimate of effectiveness) of technology use in classroom, online and blended learning as reflected in student achievement outcomes?
2. What moderator variables (methodological, substantive and contextual) influence learning with technology in these educational settings and to what extent?
3. Does the methodological quality of primary meta-analyses under review matter?
4. Does the major function (objective) of technology use affect learning outcomes?
5. Is learning differentially successful in fields of study (subject matter) when supported by educational technology?
6. Is educational technology differentially effective for learners at different academic grade levels/ages?
7. How effective was technology use for learning over time?
8. What additional aspects and emerging directions of research on technology integration in education deserve closer examination in subsequent systematic reviews?

Methodology

It is well accepted that no methodology is completely immune to bias, and like all other forms of research, meta-analyses may be biased in several ways. Considering that the aim of meta-analyses is to evaluate the state of an entire collection of related primary research, however, the consequences of bias are potentially much more detrimental than in individual primary research studies. Moreover, as rightfully noted by Bernard (2014), potential bias in systematic reviews, and especially in meta-analyses, may lead to greater distortion in the representation of research findings than bias in primary empirical studies, as errors, inconsistencies and/or omissions in the latter could be accumulated and thus amplified in the former, unless serious efforts are taken to detect and prevent bias in research syntheses. As such, sources of bias in meta-analyses have received growing attention in recent years, especially with the exponential increase in the number of published meta-analyses and systematic reviews. Bias in meta-analysis may be an intentional or unintentional systematic inaccuracy that results from the processes employed in data collection, analysis or interpretation (Bernard, 2014).

To address the issue of bias, a meta-analysis methodological reporting quality guide (MMRQG) was designed, developed and verified with a set of 52 meta-analyses addressing educational technology published between 1988 and 2017 (see below). The MMRQG is a rubric to help consumers of meta-analytical research — policymakers, practitioners and novice researchers — to avoid pitfalls associated with misinterpreting findings of a given meta-analysis (either overstating or underestimating their applied value) as a result of potential bias associated with shortcomings in the implementation of such a complex multi-component research methodology (Bernard et al., 2018; Tamim et al., 2021). It comprises 22 items (presented in Table 3) that were designed to assess 22 areas of reporting quality in meta-analyses. The items are paired with brief descriptions that enable users to score each aspect of a meta-analysis with the aim of reflecting whether and to what extent it conforms to the rigorous standards of meta-analytical research. Based on the information available in the respective reports, each item has a corresponding three-level qualitative characterisation as follows:

1. Yes: requirements for proper meta-analytical procedures/decisions are fully met
2. Somewhat: requirements are partly met
3. No/not reported: requirements are not met, or the corresponding information is missing

Once a meta-analysis is assessed by coding its quality across the 22 MMRQG items, it becomes a matter of simply summing all the codes to calculate an index of methodological quality with a theoretical range from 0 (all items coded No) to 44 (all items coded Yes). See Tamim et al. (2021) for full details about the MMRQG and its use.

Table 3. MMRQG: Items and descriptions

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

Findings from a verification study with 52 meta-analyses addressing educational technology revealed an inverse relationship between MMRQG scores and effect size. This means that lower-quality meta-analyses predict higher effect sizes than higher-quality ones. Thus, findings from less robustly conducted meta-analyses may mislead practitioners and policymakers as they tend to overestimate the effect size of the treatment under investigation. Examination of individual items of the MMRQG revealed that nine items contribute significantly to this inverse relationship:

- contextualisation of the problem
- defining the experimental group
- defining the control group
- specifying the outcome measures
- reliability of article review
- validity of included studies
- independence of data
- treatment of outliers
- moderator variable analyses

For the purpose of the current second-order meta-analysis, and considering the substantive number of published meta-analyses addressing technology integration in educational contexts, we decided to use one particular criterion that (according to findings by Tamim et al., 2021) sets apart methodologically strong meta-analyses from weaker ones: attention to moderator variable analysis. As such, to ensure more rigour and higher reliability of findings and avoid overestimation of the average effect size, we considered for inclusion only meta-analyses that carried out some form of analysis of moderator variables (methodological – e.g., research design; substantive – e.g., pedagogical framework; or demographic – e.g., age of learners).

In implementing our review, we followed the standard seven-step procedure for conducting systematic reviews as outlined in Cooper (2017), but with a focus on published relevant meta-analyses rather than primary empirical research studies. It is important to keep in mind that these steps are iterative. The implementation of a typical meta-analytical study is a non-linear process in which individual steps of the review interact with and inform each other. (For an overview of the methodology of meta-analysis, including second-order meta-analysis, see Appendix A.) The steps are outlined below.

Seven-step procedure for conducting systematic reviews

1. STATE THE REVIEW PROBLEM AND OPERATIONALLY DEFINE THE RESEARCH QUESTION(S)

State the review problem and operationally define the research question(s) and all its components (concepts and variables under consideration and the related key terminology) to enable, for example:

- a systematic literature search
- development of the criteria for selecting relevant studies for inclusion in the review
- a formalised description of experimental and control conditions
- the creation of a list of assessment tools for the dependent variable(s)
- the evaluation of methodological quality of admitted studies and data collection methods
- the creation of a preliminary list of moderating variables

In the current second-order meta-analysis, the main research question focuses on the effectiveness of educational technologies (computer-based instructional tools and applications) on students' learning outcomes (assessed by various means, including standardised performance tests, final exams and other forms of knowledge and skills evaluation) in formal educational settings without limits on grade level, subject matter or instructional delivery modes (i.e., including classroom, online and blended learning).

2. DEVELOP AND IMPLEMENT A COMPREHENSIVE SYSTEMATIC SEARCH STRATEGY

Develop and implement a comprehensive systematic search strategy (according to the keywords selected at the first stage) for relevant research literature.

We conducted a systematic search for primary (first-order) meta-analyses in the major educational electronic databases — ERIC (EBSCO), Education Source (EBSCO) and ProQuest Education Database — using comprehensive lists of controlled vocabulary descriptors available in each database to target the concept of technology, joined with the inclusion of "meta-analysis" in the title, abstract or descriptors and a third concept targeting achievement outcomes. These database searches resulted in a total of 641 publications for abstract screening after we removed duplicates and decided to limit the results to those published from 2000 onward. (For a complete record of the searches performed, see Appendix B.)

Grey literature was not targeted, as the probability of finding comprehensive reliable meta-analyses to inform our research question was considered negligible. Several publications that contained lists of previous meta-analyses in the subject area of educational technology were identified and set aside. We conducted citation searching to ensure that all the listed publications were in our own collection. This process identified an additional two studies, for a total collection of 641 publications. After screening the abstracts, we coded 214 of them for full-text retrieval. Of these, three were impossible to retrieve, leaving a final full-text collection of 211. After a thorough review of full-text documents, a final set of 131 studies (featuring 134 effect sizes) was retained for the final analysis. The results of each stage of the review are detailed in the PRISMA diagram in Figure 1 (page 23).

3. SELECT STUDIES FOR INCLUSION IN THE REVIEW AND COLLECT DATA FROM THEM

Select studies for inclusion in the review and collect data from them, including extracting effect sizes and coding moderator variables. The study selection unfolded in two stages: abstract screening followed by a full-text review of the selected documents for final inclusion in the second-order meta-analysis. To expedite the review process without compromising its reliability, the inter-rater agreement rate for the independent decision-making was established through the following sequence of actions: First, the entire collection of identified abstracts was split among the team members, all experienced researchers with at least a decade of experience in conducting systematic reviews, who screened their respective portion of studies independently. In this preliminary review they set aside abstracts that needed a second opinion so that the entire team could discuss whether or not to include them. This allowed a range of reliability measures (expressed as Cohen's kappa) to be established — from the percentage of studies (out of the initial total) that merited a second opinion to the portion whose original independent judgement was confirmed through the team discussion. The same procedure was employed for the full-text documents' review, as well as at the effect size extraction and coding of moderator variables (with several discussion sessions throughout the entire review implementation) stages. We aimed to achieve a reliability rate of around 0.7 for either of these metrics as the assurance of the overall reliability of all types of coding — low and high inference (Cooper & Hedges, 1994).

Note that this current review does not estimate the degree of potential overlap in individual primary research studies in the admitted meta-analyses. As shown by Tamim et al. (2011), in the supplementary validation study, this time- and labour-consuming complex procedure does not add precision to the overall findings of a second-order meta-analysis. In our view, the 2011 validation study established, in principle, the reliability of the second-order meta-analysis methodology.

The following inclusion/exclusion criteria were used at all stages of the review:

- A meta-analysis under consideration should be published from 2000 onward.
- Results should be reported aggregated in the form of the overall weighted average effect size (Cohen's *d*, Hedges' *g* or Pearson's *r*), reflecting assessment of the effectiveness of educational technology use in in-class, online or blended learning educational settings.
- A meta-analysis must implement some form of moderator variable analysis.
- Only quantitative syntheses (true meta-analyses) were admitted to the final analytical stage. Neither descriptive systematic reviews nor quantitative summaries (i.e., reviews that did not aggregate the overall effect size, though possibly reported effects from individual studies) were retained.
- Two broad categories of meta-analyses were excluded (though recorded for potential future use): (1) those with special populations — special needs students (e.g., diagnosed with some form of learning disability), and (2) those that compared technology use with some enhancement (typically pedagogy-based, e.g., collaborative learning) in the experimental condition and unmodified use of the same technology in the control condition (we designated those Added Value meta-analyses).
- Several second-order meta-analyses (mostly on flipped classrooms, as well as our own and Tamim et al., 2011) were also excluded.

4. ASSESS THE METHODOLOGICAL QUALITY, RELEVANCE, PRECISION AND COMPREHENSIVENESS OF STUDIES INCLUDED IN THE META-ANALYSIS

In parallel with the step above, assess the methodological quality, relevance, precision and comprehensiveness of studies included in the meta-analysis; and identify and correct research design and statistical confounds to ensure the overall reliability and validity of the resulting research synthesis.

In the current second-order meta-analysis, we relied on assessments of the methodological quality of primary (first-order) meta-analyses under review both to inform inclusion/exclusion decisions and to act as a methodological moderator variable to make sure that variation in the quality of the admitted studies did not affect the overall review findings. We used the MMRQG to implement this assessment. The MMRQG directs readers' attention to the key components of meta-analysis implementation and helps to identify and account for its strengths and weaknesses. We used MMRQG items that, in previous research, emerged as the most significant predictors of the distribution of effect sizes (e.g., adequacy and comprehensiveness of search strategy/data sources/analytical procedures, attention to quality assessment of included primary studies, and performance of moderator variable analysis).

More specifically, we coded and analysed the following moderator variables:

- **Subject matter:** We distinguished among Language Arts, Maths and Statistics, Science, and Medical Education, when specified. The vast majority of the included meta-analyses, however, dealt with either multiple subject matters or combined categories identified by their authors (e.g., four cases of STEM).
- **Grade level:** To reflect the learners' academic levels and ages, we used several categories. Whenever academic levels were reported, a distinction was made among pre-kindergarten and kindergarten, elementary school, secondary and high school, post-secondary education (including colleges) and various forms of training for adults. For analytical purposes, as some specific categories were represented by a very limited number of cases, we tried creating larger categories — e.g., collapsing the former three into a Compulsory Schooling category and the latter two into an Above Compulsory School category. Also, as with the subject categories, a high percentage of the included meta-analyses reported findings across grade levels (all-encompassing or in various combinations).
- **Publication date and coverage:** To reflect the range of time periods during which the primary studies included in a meta-analysis were conducted and how old those studies were, we split the entire collection into meta-analyses published between 2000 and 2010, inclusive, and those published from 2011 onward. We also considered the publication dates of primary empirical studies included in meta-analyses in these two categories, as some of the more recent meta-analyses went back several decades in their coverage of primary research, which created an overlap in coverage between them and older meta-analyses.
- **Major function of using technology:** To reflect the predominant objective(s)/expectations arising from employing a particular type of educational technology, we classified the included meta-analyses largely based on the categorisations developed and used in some of our previous works (e.g., Borokhovski et al., 2016; Schmid et al., 2014; Tamim et al., 2015). Specifically, we tried to distinguish among educational technologies primarily employed to provide cognitive support (both for deep learning — e.g., simulations — and for distributed cognition — e.g., SPSS), enable/enhance presentation of learning materials and/or access to information, facilitate interaction/collaboration among learners and promote learning through gaming. It is important to note, however, that the majority of the reviewed meta-analyses encompassed multiple types (and therefore functions) of technology use. Special attention was paid to methodological moderator variables. Depending on the availability of the related information in the reviewed reports, we considered:

- How thorough and complete the procedures for assessing the methodological quality of primary empirical studies within individual meta-analyses (e.g., coding and analysing research design, addressing the dependency issue) were and how closely their own implementation procedures were observed. For analyses, we created broad categories: High level/Limited/No verification of methodological quality.
- To what extent and depth the analysis of moderator variables was implemented in the reviewed meta-analyses. Specifically, we looked at whether sufficient rationale for the moderator variable analysis was provided and how many meaningful moderators were coded and explored. As a result, two broad categories were coded and analysed: Comprehensive and Partial moderator variable analyses. The complete absence of moderator variable analysis warranted the exclusion of a study.
- Whether the source/category of the achievement outcome measures was specified in different meta-analyses. Though most of the included meta-analyses listed either a high range of achievement measures or a generic "achievement"/"performance" label, some did explicitly identify what category of learning achievements was assessed (e.g., reading comprehension or vocabulary knowledge), and this factor was accounted for in our own moderator variable analyses.

Table 4. Levels of moderator variables represented in the included meta-analyses (frequencies) across all three collections (k = 134)

<table>
<tr><th colspan="5">Assessment of methodological quality of included primary studies</th></tr>
<tr><td>Multiple verifications</td><td colspan="3">Limited assessment</td><td>None</td></tr>
<tr><td>51</td><td colspan="3">59</td><td>24</td></tr>
<tr><th colspan="5">Analysis of moderator variables</th></tr>
<tr><td colspan="2">Comprehensive set of moderators</td><td colspan="3">Limited number of selected moderators</td></tr>
<tr><td colspan="2">109</td><td colspan="3">25</td></tr>
<tr><th colspan="5">Type of outcome measure/source</th></tr>
<tr><td colspan="2">Variety of generic measures</td><td colspan="3">Specific outcome measures</td></tr>
<tr><td colspan="2">87</td><td colspan="3">47</td></tr>
<tr><th colspan="5">Publication date coverage (end-date of included studies)</th></tr>
<tr><td colspan="2">Up to 2010 (inclusive)</td><td colspan="3">From 2011 onward (inclusive)</td></tr>
<tr><td colspan="2">31</td><td colspan="3">103</td></tr>
<tr><th colspan="5">Data source by region</th></tr>
<tr><td colspan="2">International coverage</td><td colspan="3">Single region data source</td></tr>
<tr><td colspan="2">118</td><td colspan="3">16</td></tr>
<tr><th colspan="5">Age/grade level of learners: Age/grade level of learners*</th></tr>
<tr><td>All grade levels/age groups</td><td>Pre-K/K & elementary</td><td>K-12 (all grades)</td><td>Mixture of secondary & post-secondary</td><td>Post-secondary & adult education</td></tr>
<tr><td>61</td><td>7</td><td>33</td><td>6</td><td>25</td></tr>
<tr><th colspan="5">Subject matter/field of study</th></tr>
<tr><td>Language arts & other non-STEM</td><td>Mathematics & statistics</td><td>Medical education</td><td>Science & non-maths STEM</td><td>Across disciplines</td></tr>
<tr><td>36 (34+2)</td><td>20</td><td>4</td><td>10</td><td>64</td></tr>
</table>

* Total number of cases for this moderator variable does not add up to 134, as there are N/A cases.

5. STATISTICALLY AGGREGATE INDIVIDUAL EFFECT SIZES TO DERIVE AN OVERALL WEIGHTED AVERAGE INDEX

Statistically aggregate individual effect sizes to derive an overall weighted average index reflecting the effectiveness of the experimental treatment (in comparison with the control one) in the entire general population; assess potential biases; and analyse moderating variables to explain the spread and systematic variations in the distribution of effect sizes. We used the Comprehensive Meta-Analysis software package (Borenstein et al., 2014), version 3.3.070, for all statistical analyses in this project.

Second-order meta-analysis presents an interesting challenge when deciding on the sample size in estimating the standard error associated with each individual effect size. This decision determines each effect size's relative contribution to the aggregated overall effect size. While in individual primary studies relative weights of individual effects are based on the inverse variance, which is the function of the corresponding sample size (i.e., number of participants in both conditions), a second-order meta-analysis should not rely on this procedure as there is no way of accounting for potentially disproportionate distribution of large and small (in the number of participants) studies across admitted first-order meta-analyses. Instead, we calculated weights for each effect size extracted from these meta-analyses based on the doubled number of independent effect sizes in every one of them (i.e., the number of independent comparisons, each necessarily involving an experimental and a control condition — hence, increased by a factor of two). We then used the random effects model for the aggregation of individual effect sizes (Borenstein et al., 2009).

Standard analytical procedures (e.g., one-study-removed, classic fail-safe, Duval and Tweedie's trim and fill, funnel plot examination) were performed to detect, and, if necessary, correct for, potential publication and outlier biases (see Borenstein et al., 2009, and Duval & Tweedie, 2000).

Coding, and more importantly analysis, of moderator variables largely depends on what categories are consistently accounted for across admitted meta-analyses. We made a serious effort to identify all repeatedly available moderator variables and use the mixed effects analytical model to explore and explain variations in the outcomes associated with the systematic influence of these variables. In addition to major categories of instructional delivery mode — classroom, online and blended learning — we hoped to find sufficient information about, at least, "methodological quality" (Bernard, Borokhovski, Schmid & Tamim, 2014), "technology type and functionality" (Schmid et al., 2014), including its interactive potential (Bernard et al., 2009; Borokhovski et al., 2012; Borokhovski et al., 2016), "grade level" and "subject matter" moderator variables.

6. INTERPRET THE RESULTS IN A WAY THAT ALIGNS WITH THE MAIN RESEARCH QUESTION(S)

Interpret the results obtained in a systematic review in a way that aligns with the main research question(s) and highlights both implications for practice (by informing and enabling effective decision-making — e.g., based on analyses of "what works?" kind of evidence) and guidance for subsequent primary empirical studies (targeted to identify, collect and expand such evidence).

In this second-order meta-analysis, we considered three independent sets of findings aggregated around three major delivery modes — classroom, online and blended learning — in the focus of our research question about the effectiveness of educational technology use. For each mode we addressed clarifying questions (based on the analyses of available moderator variables) of applied value to enable practice-oriented recommendations (i.e., "What works under what circumstances and for whom?").

7. PRESENT THE RESULTS IN A TRANSPARENT FASHION

Present the results of a systematic review (meta-analysis) in a transparent fashion with regard to the review's methodology and step-by-step implementation. It should be sufficiently detailed about its findings and place an emphasis on practical recommendations for the target audience(s).

Abstract screening and full-text review outcomes

As depicted in the PRISMA flow chart (below), we began with 915 potentially suitable records, identified through initial searches, and ultimately included only 131 in our review. At the final stage of the inclusion/exclusion review we had 211 full-text documents. We rejected 79 of them for a variety of reasons:

1. The reviewed studies, though quite broad in their coverage, were not in fact meta-analyses. Instead, they represented some form of descriptive summary of research findings (e.g., Burtson & Arispe, 2018, or Wang & Nunes, 2019), large-scale primary studies (e.g., Gunter, 2012) or comprehensive opinion papers (e.g., Simonson, 2015) — that is, they did not extract and analyse effect sizes from the reviewed primary empirical studies. Nineteen papers were excluded for this reason.

2. Some studies, named meta-analyses, did not produce any aggregated data (overall weighted average effect size), but simply presented effect sizes extracted from included primary empirical studies, individually (Zucker et al., 2009) or by category (Al-Jewair et al., 2009). Sixteen studies were excluded for this reason.

3. As outlined in the section on the MMRQG and in the inclusion/exclusion subsection, meta-analyses that did not carry out any moderator variable coding and analysis (e.g., Blok et al., 2002) were not taken into consideration. Sixteen meta-analyses were excluded for this reason.

4. Other reasons for exclusion were not repeated frequently enough to warrant the creation of independent categories for them. Some meta-analyses reported obvious untreated outliers (e.g., Al-Wasy, 2020), and several others represented either subsets or data duplicates of larger meta-analyses published elsewhere (e.g., Lou et al., 2006, appeared to be a reconfigured and reanalysed data selection from Bernard, Abrami, Lou et al., 2004). Obviously, other second-order meta-analyses (e.g., Hew et al., 2021, or Tamim et al., 2011) were excluded from consideration either at the abstract screening or the full-text document review stages. In total, this category of excluded studies comprised 29 papers.

The PRISMA diagram does not take into account meta-analyses set aside in the categories of Special Needs Students (18 effect sizes) and Added Value (26 effect sizes), which we believe deserve special consideration, and as such could not be collapsed, analysed and interpreted together with meta-analyses based on studies of general populations and focusing on the "technology versus no technology" comparison type. We briefly discuss both these collections below. A thorough examination of the growing body of systematic reviews in these research areas is beyond the scope of this current publication but merits further research in the future.

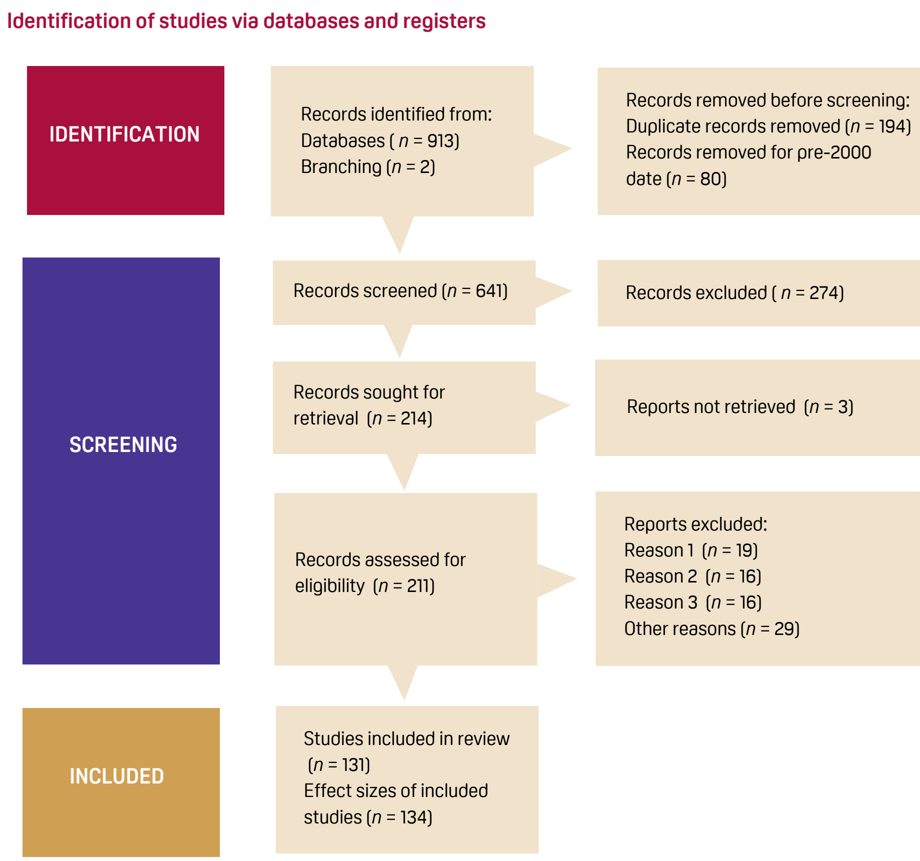

Figure 1. PRISMA diagram (based on Page et al., 2021).

Special needs students

This category of meta-analyses investigates technology use intended to facilitate learning among special populations of students. It encompasses a wide range of health conditions that generally fall into one of two categories: learning disabilities (e.g., reading difficulties) and disabilities directly linked to mental or physical health issues (e.g., autism – in four meta-analyses). For this group of students, educational technology is employed to help them learn. As indicated earlier, we located 18 such meta-analyses. As in all major collections in our second-order meta-analysis, they feature various types of technology, from narrowly focused software applications (e.g., text-to-speech and related read-aloud tools for students with reading disabilities, as in Wood et al., 2018) or video-modelling technologies for preschoolers with autism spectrum disorder (Lyngdoh & Bene, 2018) to mobile devices for intellectually impaired learners (Kim & Kimm, 2017) and more generic multifunctional computer-assisted instruction technologies (CAI) to improve performance in mathematics among students with learning disabilities (Seo &

Bryant, 2009). On average, the methodological quality of meta-analyses that formed this collection was compatible with that of the entire data set. Three meta-analyses summarised data from so-called single-case (or single-subject) studies, more than we found in any other collection of either included or set-aside meta-analyses. We are also aware of the rapidly growing literature on the educational use of artificial intelligence (AI) among special populations. At time of writing, however, no meta-analysis has been created from this literature, only descriptive reviews (e.g., Fichten et al., 2021).

Added value

Another category of meta-analyses that in our view merits examination separately from those included in this current second-order meta-analysis comprises studies that compare different educational technologies to each other according to type, degree of frequency/intensity of use and whether they are augmented with some instructional (i.e., pedagogical) or functional add-ons. The number of meta-analyses in this category is relatively low — 26 (counting all three types of educational settings) — which appears to reflect the course that educational technology primary research has taken since its early days and that currently prevails. Specifically, this most widespread approach (which dominates the research literature) focuses on comparing technology-supported (technology-enabled) instructional interventions to "technology-free" educational environments. The vast majority of research studies, and by extension research reviews (including meta-analyses), apply this "all or none" paradigm.

This happens because, to some degree, research is relatively easy to conduct when there is a natural treatment condition (the presence of something) and control condition (the absence of the same). Not only are such comparisons natural and easy to code, but possible alternatives are also difficult to conceive in an educational context, as technology presence across conditions creates variability that is very challenging to meaningfully categorise into consistent intervention and control groups. The problem is that the research literature has not kept pace with either educational practices regarding the application of technology to teaching and learning processes or educational practitioners' desire for a more nuanced (i.e., better informed by instructional design and pedagogical frameworks) use of technology. It is probably quite difficult now to find classrooms at any level of the education system that are completely technology-free.

Shifting the focus of research on educational technology away from oversimplified "yes/no" comparisons to more informative research that is representative of real classroom comparisons between different instructional uses of technology (e.g., Bernard et al., 2009, and Borokhovski et al., 2012, for distance education; or Schmid et al., 2014, for post-secondary education with technology) is referred to in this current report as "added value" — that is, sound pedagogy, instructional design and/or enhanced technological functionality in the experimental condition are supposed to add some learning benefits to those derived from generic (unmodified/unguided) use of the same technology in the control condition.

The methodological quality of the “added value” meta-analyses is comparable to that of the retained collections. Even more important, some of those meta-analyses represent a subset of the retained collections (i.e., some meta-analyses not only reference the “technology+ versus technology alone” type of comparisons, but also analyse and report them separately). Also, they appear to be more recent: 22 were published after 2010 and only four prior to 2010. This may be indicative, albeit indirectly, of the fact that attention to added value research gradually intensifies as researchers strive to better understand the topic and that the added value aspect therefore merits investigation in the future. For the purposes of the current report we simply summarise some preliminary observations.

Unlike the retained collection, the Added Value meta-analyses represent only two types of educational settings: in-class technology integration (e.g., van der Kleij et al., 2015) and online learning (e.g., Bernard et al., 2009). They do not represent blended learning — perhaps because this learning mode is itself often perceived as an improvement on classroom and distance education by virtue of combining the two. The technologies they feature range from generic CAI or multimedia use to the targeted use of such sophisticated tools as computer simulations, digital game-based learning (DGBL) and augmented (virtual) reality (AR/VR). Their major focus is on how and how much educational technology can be enhanced by the added value that comes with specific pedagogy and instructional design, such as scaffolding (Doo et al., 2020); designed collaborative interactions (Borokhovski et al., 2016); self-regulated learning (Zheng, 2016); metacognitive instruction (Lan et al., 2014); strategic discussions versus conventional online learning (Darabi et al., 2013); the competition factor in educational games (Chen, Shih et al., 2020); and others. Similarly, they explore the extra functionality of technology itself — e.g., embodied pedagogical agents (Guo & Goh, 2015), dynamic (versus static) visualisations in science instruction (McElhaney et al., 2015) or automated adaptive guidance in CAI (Gerard et al., 2015). Meta-analyses in this collection were also all-encompassing in terms of the covered fields of study and grade levels.

Results

As depicted by the PRISMA flow diagram (Figure 1, page 23), overall, 131 published meta-analytical investigations featuring 134 individual effect sizes were selected for inclusion in this second-order meta-analysis. They were distributed among the three categories of interest:

- In-class technology integration: 108 effect sizes
- Online learning (technology-enabled): 12 effect sizes
- Blended learning (technology-supported): 14 effect sizes

Inter-rater agreement rates, estimated as outlined in the Methodology section, were as follows: The team review of the abstracts set aside as questionable in the first round of independent screening (60 out of 719, or 8.34%) confirmed 43 (94.02%) initial decisions. These numbers translate into a Cohen’s kappa of 0.88. The same procedure carried out at the stage of full-text review resulted in a Cohen's kappa of 0.65 (14 out of 17 confirmed initial decisions). This relatively sizeable drop in reliability could be explained by the fact that the status of some categories of the reviewed meta-analyses (e.g., the Added Value collection of meta-analyses) was not predetermined but was instead decided upon through team discussions during the project implementation.

As stated in the Methodology section, all types of the ES metric (i.e., Cohen's d, $k = 50$; Hedges' g, $k = 76$; and Pearson's r, $k = 2$) were admitted. In the two latter cases, the correlation coefficient r was converted into Cohen's d as follows: $d = 2r / \sqrt{1 - r^2}$. Whenever the reports did not specify the ES metric ($k = 6$), it was assumed that the aggregated ES was expressed as the d-type. Subsequently, all d effect sizes were converted into the common g metric for further analyses and aggregation.

The upcoming analyses (including publication bias, sensitivity and methodological, substantive and contextual moderator variables) will be presented according to standard meta-analytical procedures to estimate the effects of technology use on learning achievement outcomes in classroom, online and blended educational settings and to explore and, when possible, explain any variability that surrounds the corresponding overall weighted average effect sizes. We created three tables — one for each type of educational setting — to present some key aspects of all the included meta-analyses in more detail.

Column 1 of each table contains each meta-analysis ID (authors and publication date). Column 2 contains the magnitude of the overall effect size, as reported in the study, and the number of cases it is based on (individual comparisons in the included primary empirical studies). Column 3 contains information about the sample (total number of students, when reported, and academic level/age of learners) and the learning context (primarily, subject matter or field of study with some additional details, when important). Column 4 provides a brief description of technology type and functionality used in the corresponding meta-analysis. Column 5 offers a succinct summary of each meta-analysis in terms of focus and principle message.

Table 5. Key descriptors of meta-analyses in the In-Class Technology Integration collection (k = 94)

Study ID	Reported effect size (ES)	Sample & context characteristics	Technology type & functionality	Review focus & summary of findings
Larwin & Larwin (2011)	0.566 (k = 219)	40,125 post-secondary students learning statistics	Generic computer-assisted instruction (CAI), multipurpose use	A meta-analysis (MA) synthesising 219 achievement ESs from 70 studies published between 1970 and 2010, addressing the effectiveness of CAI on student achievement in post-secondary statistics education. Findings revealed a moderate positive overall ES in favour of the CAI. Follow-up analyses indicated that CAI is most beneficial when used as a supplementary learning source, for computation purposes and for simulations.
Grgurovic et al. (2013)	0.235 (k = 49)	Unspecified number of language learners across grade levels	Tests the effects of multipurpose use of computer-assisted language learning (CALL) on various language skills	An MA synthesising 49 achievement ESs from 37 studies published between 1984 and 2006, addressing pedagogies for second/foreign language teaching supported with computer technology. A small but positive and statistically significant mean ES in favour of technology-supported pedagogy was found.
Yammine & Violato (2015)	0.297 (k = 36)	2,128 medical (post-secondary) students studying anatomy	Three-dimensional visualisation technologies used largely for cognitive support of learning	An MA synthesising 36 achievement ESs from 36 studies published between 1997 and 2013, addressing the effectiveness of three- dimensional visualisation technology (3DVT) in teaching and learning anatomy. Findings show a small mean ES.
Fang et al. (2019)	0.100 (k = 24)	4,116 secondary and post-secondary students studying mathematics	Multipurpose use of Assessment and Learning in Knowledge Spaces (ALEKS) intelligent tutoring systems (ITS)	An MA synthesising 24 achievement ESs from 15 studies published between 2005 and 2015, addressing the effectiveness of the ALEKS on learning. Results revealed that ALEKS was as good as, but not better than, traditional classroom teaching. Moderator analyses revealed the ESs were not significantly different when ALEKS was used by students at different schooling levels, when it was implemented differently or when the learning outcomes were measured differently. ESs were greater when ALEKS was used for shorter periods rather than longer periods.

Bayraktar (2002)	0.272 (k = 108)	Unspecified number of secondary and college students in science education	Various forms of CAI: drill & practice, simulations, tutorials	An MA synthesising 108 achievement ESs from 42 studies published between 1970 and 1999, addressing the effectiveness of CAI on student achievement in secondary and college science education. Findings revealed a weak mean ES, suggesting that a typical student moved from the 50th percentile to the 62nd in science when CAI was used. Student-to-computer ratio, CAI mode and duration of treatment were significantly related to the effectiveness of CAI.
Li & Ma (2010)	0.279 (k = 85)	36,793 K–12 students studying mathematics	Generic CAI multipurpose use	An MA synthesising 85 achievement ESs from 46 studies published between 1990 and 2006, addressing the impact of computer technology (CT) on mathematics education in K–12 classrooms. Findings indicated statistically significant positive effects of CT in promoting the mathematics achievement of elementary over secondary school students, and CT showed larger ESs on the mathematics achievement of special needs students than that of general education students.
Ran et al. (2021)	0.555 (k = 45)	2,044 K–12 students studying mathematics	Tests effects of multipurpose CAI (tutorials, gaming, drill & practice, problem-solving) on maths understanding, reasoning, etc.	An MA synthesising 45 achievement ESs from 31 studies published between 2003 and 2020, addressing the effects of CT on mathematics achievement, with a particular focus on low-performing students. Findings revealed a statistically significant and positive effect of CT on low-performing students' mathematics achievement. Of four CT types, the largest ES was found with problem-solving applications, followed by tutoring, game-based interventions and computerised practice.
Christmann & Badgett (2003)	0.340 (k = 68)	8,274 primary school & kindergarten students across subject matters	Unspecified CAI supplement for elementary education	An MA synthesising 68 achievement ESs from studies published between 1969 and 1998, addressing the effectiveness of traditional instruction supplemented with CAI on the academic achievement of elementary students. Findings revealed a moderate mean ES, indicating that students receiving traditional instruction supplemented with CAI attained higher academic achievements than those receiving only traditional instruction.
Lin (2015)	0.441 (k = 59)	Estimated over 3,500 second language learners across grade levels and age groups	Use of computer-mediated communication (CMC) for language learning (including elements of online learning)	An MA synthesising 59 achievement ESs from 59 studies published between 2000 and 2012, addressing the effectiveness of CMC in second language acquisition (SLA). A positive medium overall effect for CMC used for instructional/learning purposes was found.

Belland et al. (2015)	0.530 (k = 17)	Unspecified number of students across all grade levels and age groups in science, technology, engineering and mathematics (STEM) disciplines	Explore effects of computer-based scaffolding as cognitive support for learning on learners' cognitive outcomes	An MA synthesising 17 achievement ESs from studies addressing the influence of computer-based scaffolding characteristics and study and test score quality on students' cognitive outcomes in STEM disciplines. Results indicated a positive moderate mean ES. There were no differences based on study design, paired intervention, assessment level or intended learning outcome.
Torgerson & Elbourne (2002)	0.346 (k = 7)	240 primary (elementary school) students in language (specifically, spelling) training	CAI: specific software for spelling training (cognitive support: distributed cognition)	An MA synthesising 7 achievement ESs from studies published between 1980 and 2000, addressing the effectiveness of information and communications technology (ICT) on spelling. Findings revealed a non-significant mean ES in favour of computer interventions. Sensitivity and subgroup analyses of the results did not materially alter findings.
Santos et al. (2014)	0.539 (k = 11)	Unspecified number of students across all grade levels and subject matters	Augmented (virtual) reality (AR/VR) simulations (cognitive support for learning)	An MA synthesising 11 achievement ESs from studies published between 2006 and 2012, addressing the effectiveness of AR on K–12 students' learning. Findings revealed a moderate positive mean ES.
Wilson et al. (2019)	0.578 (k = 19)	2,570 medical school students in anatomy classes	Generic multipurpose CAI	An MA synthesising 19 achievement ESs from studies published between 1999 and 2016, addressing the effectiveness of student-centred pedagogies and CAI in increasing student knowledge gains in anatomy. Findings revealed a moderate mean ES with learners exposed to SCL and supplemental CAI outperforming their more traditionally trained peers.
Timmerman & Kruepke (2006)	0.294 (k = 118)	12,398 post-secondary (undergraduate & graduate) students across disciplines	Generic multipurpose/ modality CAI compared to several types of control conditions	An MA synthesising 118 achievement ESs from studies published between 1985 and 2004, addressing the effect of CAI on college student performance. Findings indicate that student performance gains are larger for CAI than traditional instruction, with CAI benefits being greatest for social science disciplines, when the traditional instruction format is lecture/discussion, for undergraduate samples, in studies published after 1994, and for CAI delivered in multiple units.

Chiu (2013)	0.726 (k = 16)	1,684 second language learners of all grade levels and age groups	Explores the effects of several types of CALL (e.g., presentations, language games) on vocabulary knowledge	An MA synthesising 16 achievement ESs from studies published between 2005 and 2011, addressing the general effectiveness of L2 computer-assisted vocabulary instruction, with analysis of the features of treatment duration, educational level, the use of games and the role of teachers in the CALL studies. Findings revealed a strong positive mean ES.
McTigue et al. (2020)	-0.020 (k = 19)	2,373 preschool, elementary & secondary school students receiving reading instruction	GraphoGame (GG): adaptive serious game designed to prevent reading difficulties	An MA synthesising 19 achievement ESs from studies published between 2005 and 2018, addressing the effects of GraphoGame (GG), an adaptive serious game, on students' reading skill, and specifically on word-reading outcomes. Findings did not yield an overall meaningful ES.
Guo et al. (2020)	0.390 (k = 39)	2,103 second language learners of all grade levels and age groups	Specific CAI application for including graphics in instruction for reading comprehension	An MA synthesising 39 achievement ESs from 39 studies published between 1985 and 2018, addressing the effectiveness of well-designed graphics on readers' understanding of a text. Findings revealed that the inclusion of graphics has a moderate overall positive ES on students' reading comprehension, regardless of grade level.
Cromley et al. (2020)	0.690 (k = 166)	8,111 science learners across all grade levels	Specific CAI application: picture-drawing software for cognitive support for learning science	An MA synthesising 166 achievement ESs from 53 studies published between 2005 and 2018, addressing the effect of directed learner-generated visual representations on students' learning in science. Findings revealed a significant positive mean ES for drawing-to-learn across all dependent variable types (factual, inferential and transfer) and both types of effects.
Chan & Leung (2014)	0.971 (k = 9)	587 K–12 students studying mathematics (geometry)	Specific CAI application: dynamic geometry software used for presentation and production	An MA synthesising 9 achievement ESs from 9 studies published between 2002 and 2012, addressing the effectiveness of dynamic geometry software (DGS) in improving students' mathematical achievement. The overall ES of DGS-based instruction on achievement scores was strong. Subgroup analysis found some groups benefited more than others — for example, short-term instruction with DGS significantly improved the mathematical achievement of elementary school students.
Sosa et al. (2011)	0.327 (k = 45)	9,639 post-secondary students in statistics courses	Specific CAI software (e.g., SPSS) for cognitive support for learning (distributed cognition)	An MA synthesising 45 achievement ESs from 45 studies published between 1974 and 2005, addressing a range of specific features that presumably influence the effectiveness of CAI. Findings revealed a moderate positive mean ES. Analyses revealed that larger effects were reported in studies in which treatment groups received more instructional time than control groups, in studies that recruited graduate students as participants, and in studies employing an embedded assessment.

Kim et al. (2018)	0.385 (k = 47)	Unspecified number of predominantly secondary and post-secondary students in STEM education	Specific CAI application: computer-based scaffolding for cognitive support for learning	An MA synthesising 47 achievement ESs from studies published between 1990 and 2015, addressing the effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education. Results indicated that computer-based scaffolding significantly positively impacted cognitive outcomes in problem-based learning in STEM education.
Kulik & Fletcher (2016)	0.660 (k = 50)	20,443 students across grade levels and subject matters	Specific CAI application: multipurpose ITS	An MA synthesising 50 achievement ESs from 50 studies published between 1993 and 2013, addressing ITSs. Findings revealed a medium-strength positive mean ES.
Mahdi (2018)	0.660 (k = 16)	986 adult second language learners	Use of various mobile devices (MALL) to support language (vocabulary) acquisition	An MA synthesising 16 achievement ESs from 16 studies published between 2009 and 2015, addressing the effectiveness of using mobile devices for vocabulary learning. The findings of the meta-analysis indicate a medium effect of using mobile devices on vocabulary learning. In addition, it was found that adult learners benefit more than young learners from using mobile devices for vocabulary learning.
Tokac et al. (2019)	0.129 (k = 39)	Unspecified number of preschool and compulsory (K–12) students studying mathematics	Digital game-based learning (DGBL): use of video games for learning	An MA synthesising 39 achievement ESs from 24 studies published between 2000 and 2017, addressing the effects of learning video games on mathematics achievement of pre-K–12 students compared with traditional classroom instructional methods. Findings suggested that mathematics video games contributed to higher learning gains.
Holmes (2013)	0.214 (k = 14)	1,126 preschool and compulsory (K–12) students studying mathematics	CAI: use of virtual manipulatives for cognitive support of learning	An MA synthesising 14 achievement ESs from studies published between 1995 and 2012, addressing the use of virtual manipulatives for cognitive support with K–12 maths teaching and learning. Findings indicated a small positive mean ES.

Tingir et al. (2017)	0.483 (k = 14)	Unspecified number of K–12 students across disciplines	Multipurpose use of various mobile devices (iPads, tablets, smartphones, personal digital assistants [PDAs])	An MA synthesising 14 achievement ESs from 14 studies published between 2010 and 2014, addressing the effects of mobile devices on student achievement in science, mathematics and reading in K–12 students. Results suggest that using mobile devices in teaching yielded higher achievement scores than traditional teaching in all subject areas.
Higgins et al. (2019)	0.675 (k = 52)	4,522 mathematics students from preschool to Grade 8	Generic CAI multipurpose use	An MA synthesising 52 achievement ESs from 24 studies published between 1985 and 2013, addressing the effects of generic technology use on student achievement. Results indicated a significant overall impact of technology on student achievement.
Saad Mohamed (2020)	0.560 (k = 21)	1,313 second language learners in post-secondary (undergraduate & graduate) education	Tests effects of feedback in the CALL environments for cognitive support of learning	An MA synthesising 21 achievement ESs from 21 studies published between 1992 and 2016, addressing the use of feedback in CALL. Findings indicated that feedback in CALL has a significant medium ES on student language learning outcomes. Results also indicated that the effect of feedback is moderated by a host of variables, including learners' mother tongue, intervention provider (e.g., teacher, researcher), target language, etc.
Radkowitschet al. (2020)	0.240 (k = 54)	5,616 students across grade levels and subject matters	Computer-supported collaborative learning (CSCL) to enable and support interactions among learners	An MA synthesising 54 achievement ESs from 53 studies published between 2000 and 2020, addressing the effect of learning with a CSCL script provided to support unguided collaborative learning on students' learning. Results indicated that learning with CSCL scripts leads to a small positive effect on domain learning and that scaffolding single collaborative activities and a combination of collaborative activities affects the effectiveness of CSCL scripts.
Schroeder et al. (2013)	0.190 (k = 43)	3,088 students across grade levels and subject matters	Specific CAI application: multipurpose use of pedagogical agents	An MA synthesising 43 achievement ESs from 43 studies published between 1998 and 2011, addressing the effect of using pedagogical agents on learning. The overall mean ES was small and moderated by the contextual and methodological features of the studies. Findings revealed that the use of pedagogical agents was more beneficial for K–12 students than for post-secondary students. Pedagogical agents that communicated with students using on-screen text facilitated learning more effectively than agents that communicated using narration.
Cheung & Slavin (2012)	0.159 (k = 84)	60,553 K–12 students in reading instruction	Multipurpose generic CAI use	An MA synthesising 84 achievement ESs from 84 studies published between 1980 and 2010, addressing the effect of various technologies on K–12 students' reading outcomes. Findings suggest that educational technology applications generally produced a positive, though small, effect in comparison to traditional methods. Innovative technology applications and integrated literacy interventions with the support of extensive professional development showed more promising evidence.

Taj et al. (2016)	0.775 (k = 13)	813 second language learners across grade levels and age groups	Multipurpose use of a variety of mobile devices (PDAs, phones, etc.) and applications	An MA synthesising 13 achievement ESs from 13 studies published between 2008 and 2015, addressing mobile assisted language learning (MALL). Findings of the analysis suggest that MALL has given English as a foreign language (EFL) instruction a large positive mean ES.
Berkeley et al. (2015)	-0.030 (k = 27)	16,513 K–12 students receiving instruction in reading	Specific CAI application: using digital text in reading instruction	An MA synthesising 27 achievement ESs from 27 studies published between 2001 and 2013, addressing the impact of digital text interventions on students' comprehension when reading printed text. The overall weighted mean ES was nearly zero, whereas a moderate ES was obtained for interventions featuring instructional enhancements for digital text.
Little et al. (2018)	0.280 (k = 11)	1,528 secondary (high school) students receiving instruction in writing	Specific CAI application: CALL use to improve writing	An MA synthesising 11 achievement ESs from 6 studies published between 2003 and 2011, addressing the effect of technology-based writing instruction on writing outcomes. Results revealed a medium-strength mean ES for technology-based writing instruction. Several moderators were included in this meta-analysis, with only learning disability (LD) seeming to be a significant moderator.
Zheng et al. (2016)	0.159 (k = 67)	Unspecified number of K–12 students across subject matters	Explores the effectiveness of multifunctional "mobile" learning (one-on-one laptop computers)	An MA synthesising 67 achievement ESs from studies published between 2001 and 2015, examining the effect of one-to-one laptop programmes on teaching and learning in K–12 schools. Findings indicated a positive small mean ES.
Lin & Lin (2019)	1.005 (k = 33)	Unspecified number of second language learners of all grade levels and age groups	Specific use of MALL (text messaging, flashcards, games) for second language vocabulary learning	An MA synthesising 33 ESs from 33 studies published between 2005 and 2018, examining the connection between the use of mobile technologies and L2 word retention. Results revealed a positive and large mean ES of mobile assisted L2 word-learning interventions. Analyses also indicated that research settings, treatment durations and task-afforded autonomy are significant moderators.
Wang et al. (2020)	0.822 (k = 31)	287 students of all grade levels in language arts (L1 & L2)	Simulated virtual reality (3D virtual worlds) use for cognitive support of learning	An MA synthesising 31 achievement ESs from studies published between 2008 and 2019, addressing the impact of 3DVWs on language learning. Findings revealed a strong positive mean ES, indicating significant overall linguistic gains.

Costa & Miranda (2017)	0.498 (k = 6)	464 high school and post-secondary (college) STEM (programming) students	Specific CAI application: Alice software use for programming training	An MA synthesising 6 achievement ESs from 6 studies published between 2000 and 2014, addressing the effectiveness of the use of Alice software in programming learning when compared to the use of a conventional programming language. Findings indicated a positive medium-strength mean ES.
Sharifi et al. (2018)	0.500 (k = 158)	11,597 students of all grade levels in language arts (L1 & L2)	Multifunctional CALL use (including Web-based applications)	An MA synthesising 158 achievement ESs from 158 studies published between 1990 and 2016, addressing the effect of computer-assisted English language learning in comparison with traditional face-to-face treatments. Results indicated that CAI had an overall medium effect on English language development. Moreover, moderator analyses indicated that Web-based instruction yielded a larger mean effect than traditional CAI, with four moderators of English language learning: type of interaction, communication mode, learning context and treatment duration.
Belland et al. (2017)	0.460 (k = 333)	Unspecified number of students in STEM education across grade levels	Specific CAI application: computer-based scaffolding for cognitive support of learning	An MA synthesising 333 achievement ESs from 144 studies published between 1993 and 2014, addressing the effects of computer-based scaffolding designed to assist the full range of STEM learners from primary through adult education. Results indicated a medium positive mean ES across various contexts of use. Scaffolding's influence on cognitive outcomes did not vary on the basis of context-specificity, presence or absence of scaffolding change, and logic by which scaffolding change is implemented. Scaffolding's influence was greatest when measured at the principles level and among adult learners.
Christmann & Badgett (2000)	0.125 (k = 26)	2,032 post-secondary (college) students across disciplines	Multipurpose generic CAI use	An MA synthesising 26 achievement ESs from studies published between 1983 and 1996, addressing the effect of generic CAI on post-secondary students' learning. Findings revealed a small positive mean ES, indicating that, on average, college-level students receiving traditional instruction supplemented with CAI attained higher academic achievements than did those receiving only traditional instruction.
Özdemir et al. (2018)	0.504 (k = 16)	941 students across grade levels and subject matters	Specific CAI application: AR for cognitive support of learning	An MA synthesising 16 achievement ESs from 16 studies published between 2007 and 2017, addressing the effect of AR applications in the learning process. Findings indicated that AR applications increase students' academic achievement in the learning process compared to traditional methods. When assessing AR display devices, the largest ES was related to the use of mobile devices, while the smallest ES was in the use of webcam-based devices.
Cheung & Slavin (2011)	0.159 (k = 85)	Estimated over 60,500 K–12 students receiving instruction in reading	Multipurpose generic CAI use	An MA synthesising 85 achievement ESs from 85 studies published between 1970 and 2010, addressing the effects of technology use on reading achievement in K–12 classrooms. Findings suggest that education technology generally produced a positive, though small, effect in comparison to traditional methods. Significant moderators included education technology type, with innovative technology applications and integrated literacy interventions with the support of extensive professional development showing somewhat promising evidence.

Cheung & Slavin (2013)	0.149 (k = 74)	Estimated over 56,500 K–12 students receiving instruction in mathematics	Multipurpose generic CAI use	An MA synthesising 74 achievement ESs from 74 studies published between 1960 and 2011, addressing the effects of educational technology applications on mathematics achievement in K–12 classrooms. Findings suggest that educational technology applications generally produced a positive, though modest, effect in comparison to traditional methods, with supplemental CAI having the largest effect.
Xu, Banerjee et al. (2019)	1.280 (k = 21)	959 adult second language learners in training to improve writing quality	Several CALL applications (from Word to wikis)	An MA synthesising 21 achievement ESs from 16 studies published between 2000 and 2017, addressing the effectiveness of educational technology applications on the writing quality of adult English language learners. Findings indicated that technology applications produce a large ES compared to non-technology instructional methods. Subgroup analyses revealed that genre moderates learners' writing quality greatly, and non-collaborative technologies have a bigger impact on writing than collaborative technology applications.
Abrami et al. (2020)	0.200 (k = 91)	7,388 kindergarten & elementary school students receiving instruction in reading	Specific CAI application: multipurpose ABRACADABRA (ABRA) reading software	An MA synthesising 91 achievement ESs from 17 studies published between 2008 and 2017, addressing the effectiveness of an evidence-based suite of interactive multimedia that engages learners in the development of core reading skills (ABRA). A small positive mean ES was found. Analyses also indicated that ABRA benefits boys and girls equally and that low-performing and struggling readers were often able to learn the most and retain what was learned beyond the initial intervention.
Yun (2011)	0.460 (k = 35)	1,518 second language learners of all grade levels and age groups	Specific CALL applications: use of hypertext glosses for L2 vocabulary learning	An MA synthesising 35 achievement ESs from studies published between 1998 and 2007, addressing the effects of hypertext gloss use on L2 vocabulary acquisition in computerised reading contexts. Findings revealed a moderate positive mean ES. Further analysis indicated that studies with large samples provided a bigger ES, and that learner proficiency was found to be a statistically significant moderator.
Sung et al. (2017)	0.551 (k = 39)	5,294 learners of all grade levels across disciplines (including professional development courses)	Specific CAI application: CSCL to enable and support interactions	An MA synthesising 39 achievement ESs from studies published between 2000 and 2016, addressing the effect of mobile computer-supported collaborative learning (MCSCL). Findings revealed that MCSCL produced meaningful improvements for collaborative learning, with a moderate positive mean ES. Further analysis indicated that subject matter, group size, teaching method, intervention duration and reward method substantially moderated the ES.

Lin et al. (2013)	0.330 (k = 10)	552 second language learners of all grade levels and age groups	Specific CAI application: computer-mediated communication to enable and support interactions	An MA synthesising 10 achievement ESs from studies published between 2002 and 2012, addressing the effect of text-based synchronous computer-mediated communication (SCMC) on SLA. Findings revealed a small positive overall effect, indicating that text-based SCMC could make a larger difference on SLA than other means of communication. Findings also suggested that intermediate learners may benefit more from SCMC tasks if they work in pairs or small groups and participate in weekly SCMC interactions.
Sokolowski et al. (2015)	0.590 (k = 24)	4,256 learners of mathematics in Grades 1–8	Specific CAI application: multipurpose use of exploratory computerised environments	An MA synthesising 24 achievement ESs from 24 studies published between 2000 and 2013, examining the ES statistic of using exploratory computerised environments (ECE) to support the process of word problem solving and exploration in Grades 1–8 mathematics. Findings showed that exploratory computerised environments produced a moderate ES when compared to traditional methods of instruction. Moderator analysis revealed differences in student achievement ESs between traditional problem-solving approaches and ECEs, favouring the latter.
Chen, Wang et al. (2018)	0.450 (k = 64)	11,286 students across grade levels and subject matters	Specific CAI application: CSCL to enable and support interactions	An MA synthesising 64 achievement ESs from studies published between 2000 and 2016, addressing the effects of CSCL. Findings revealed that CSCL had significant moderate positive effects on knowledge gain.
Zheng, Zhang et al. (2020)	0.576 (k = 19)	Unspecified number of students across grade levels and subject matters	Specific CAI application: technology-facilitated peer assessment treatments	An MA synthesising 19 ESs from studies published between 1999 and 2018, addressing the impact of technology-facilitated peer assessment on students' learning. Results indicated that technology-facilitated peer assessment had a significant and medium effect on learning achievements. In addition, the use of extra supporting strategies in technology-facilitated peer assessment produced a positive and medium effect on students' learning achievements. Further analyses indicated that task types, assessment modes, training for assessors, durations, grouping types and assessment methods moderated the ES.
Noetel et al. (2021)	0.280 (k = 166)	7,776 post-secondary students across subject matters	Specific CAI application: use of asynchronous multimedia (video) for content presentation	An MA synthesising 166 ESs from 105 studies published between 1972 and 2019, addressing the effects of video on learning in higher education. Findings revealed that swapping video presentations for existing teaching methods led to small improvements in student learning. However, adding video to existing teaching methods led to strong learning benefits.
Rahmati et al. (2021)	0.904 (k = 60)	Unspecified number of language learners (L1 & L2)	Multipurpose CALL use	An MA synthesising 60 achievement ESs from 60 studies published between 2009 and 2020, addressing the impact of educational technology in English language teaching. Findings revealed a strong positive mean ES, with moderator analyses not revealing any significant moderators.

Sun et al. (2021)	0.050 (k = 56)	9,238 learners of mathematics of all grade levels	Specific CAI application: ITS (ALEKS) software	An MA synthesising 56 achievement ESs from 33 studies published between 2000 and 2020, addressing the impact of the ALEKS software on students' mathematics learning. Results revealed a null overall ES, indicating the absence of a relationship with some indication that the ALEKS was more effective when used to supplement traditional instruction than by itself.
Chen, Chen et al. (2020)	0.722 (k = 84)	Unspecified number of language learners (L1 & L2) of all grade levels and age groups	Multipurpose use of MALL (various mobile devices)	An MA synthesising 84 ESs from 80 studies published between 2008 and 2018, addressing the effectiveness of MALL. Findings revealed a medium- to high-mean ES in favour of MALL, with moderator analyses indicating that among the 9 moderator variables addressed, target language skill, target language and first/second language were found to be significant moderators.
Major et al. (2021)	0.175 (k = 15)	53,029 elementary and secondary school students in mathematics and literacy	Specific CAI application: technology-supported personalised learning in low- and middle-income countries	An MA synthesising 15 achievement ESs from studies published between 2007 and 2020, addressing the impact of students' use of technology that personalises and adapts to learning levels in low- and middle-income countries. Findings indicated that the use of technology-supported personalised learning statistically resulted in a small positive mean ES. Meta-regression revealed that more personalised approaches that adapt to learners' levels had significantly greater positive impact than those only linking to learners' interests or providing personalised feedback, support and/or assessment.
Wu et al. (2020)	0.236 (k = 35)	1,847 students across grade levels in various disciplines (with emphasis on medical and physical education, safety training, etc.)	Specific CAI application: VR simulations for cognitive support for learning	An MA synthesising 35 ESs from 35 studies published between 2013 and 2019, investigating the impact on education of immersive virtual reality (IVR) by using head-mounted displays (HMD). Results showed that IVR using HMDs is more effective than non-immersive learning approaches with a small mean ES. Moderator analyses indicated that HMDs have a greater impact on K–12 learners than on adults, in science education and specific abilities development, when offering simulation or virtual world representations and compared with lectures or real-world practices.
Bogomolova et al. (2021)	0.501 (k = 8)	3,934 post-secondary students studying anatomy	Specialised software for stereoscopic 3D visualisations for cognitive support for learning	An MA synthesising 8 ESs from studies published between 2016 and 2019, addressing the effectiveness of three-dimensional visualisation technology. Findings revealed a moderate mean ES in favour of three-dimensional visualisation technology for the acquisition of anatomical knowledge.

Howard et al. (2021)	0.540 (k = 154)	8,719 of all grade levels and age groups receiving training in special skills	Specific CAI application: VR training programmes for cognitive support of learning	An MA synthesising 154 achievement ESs from studies published between 2010 and 2020, addressing the effectiveness of VR training programmes. Findings revealed a moderate mean ES in support of VR training programmes. The results also indicated that task-technology fit and aspects of the research design moderated the effects.
Xu, Wijekumar et al. (2019)	0.600 (k = 88)	9,977 K–12 students in language arts courses	Tests the effect of using ITSs to improve reading comprehension	An MA synthesising 88 ESs from 19 studies published between 2000 and 2017, addressing the effectiveness of ITSs in improving reading comprehension for students in K–12 classrooms. Findings indicated a medium mean ES with follow-up analyses indicating a larger ES on reading comprehension when compared to traditional instruction.
Tsai & Tsai (2020), sub-collection one	0.646 (k = 14)	1,235 K–12 students in science education	Specific CAI application: using DGBL (software only)	An MA synthesising 14 achievement ESs from studies published between 2009 and 2018, addressing the effectiveness of game-based science learning against other instructional methods. Findings revealed an overall medium ES for game-mechanism design. Further subgroup analyses suggest that students across educational levels all significantly benefit from game-based science learning, although there is no significant difference between the subgroup mean effects.
Tsai & Tsai (2020), sub-collection two	0.270 (k = 13)	868 K–12 students in science education	Specific CAI application: using DGBL with mechanical components	An MA synthesising 13 achievement ESs from studies published between 2009 and 2018, addressing the effectiveness of game-based science learning when enriched with mechanisms (game-mechanism design). Findings revealed an overall small to medium ES for gameplay design with mechanical components. In addition, learning and gaming mechanisms play equal roles in significantly increasing students' scientific knowledge gains.
Petersen- Brown et al. (2019)	0.737 (k = 137)	4,017 K–12 students across subject matters	Specific CAI application: multipurpose use of touchscreen devices	An MA synthesising 137 ESs from 65 studies published between 2010 and 2018, addressing the effects of touchscreen device implementation on academic achievement. Findings revealed a positive strong mean ES with follow-up analyses indicating that guided instruction was more effective than having students work independently, and that interventions were more effective when delivered individually than in groups.
Byun & Joung (2018)	0.366 (k = 33)	2,008 K–12 students studying mathematics	Specific CAI application: multipurpose DGBL use	An MA synthesising 33 achievement ESs from 33 studies published between 2000 and 2014, addressing DGBL with K–12 mathematics learning. This study contributes to the research by analysing recent trends in significant DGBL research, especially for those who are interested in using DGBL for maths education.
Benavides-Varela et al. (2020)	0.535 (k = 15)	1,073 students in settings from pre-K to grade 12 studying mathematics	Multipurpose generic CAI (games, tutorials, drill & practice)	An MA synthesising 15 ESs from 15 studies published between 2003 and 2019, addressing the effectiveness of digital-based interventions for students with mathematical learning difficulties. Findings revealed a moderate positive mean ES with no evidence that video games offer additional advantages with respect to digital-based drilling and tutoring approaches. Effects were not moderated by grade level.

Tsai & Tsai (2018)	0.970 (k = 10)	642 second language learners of all age groups in formal and informal settings	Specific CAI application: using DGBL in L2 vocabulary acquisition	An MA synthesising 10 ESs from studies published between 2001 and 2017, addressing the effectiveness of digital games for second language vocabulary learning. Findings revealed a large overall ES that is moderated by game design, educational level, L2 proficiency level, linguistic distance, intervention setting, assessment type, game source and intervention length.
Rakes et al. (2020)	0.113 (k = 123)	Unspecified number of students across grade levels studying mathematics	Multipurpose generic CAI use	An MA synthesising 123 achievement ESs from studies published between 2001 and 2017, addressing technology use in mathematics education. Findings revealed a small positive mean ES.
Chauhan (2017)	0.545 (k = 212)	32,096 elementary school students and kindergarteners across disciplines	Tests the effectiveness of multipurpose generic CAI applications	An MA synthesising 212 ESs from 122 studies published between 2000 and 2016, addressing the impact of technology on learning effectiveness of elementary students. Findings revealed a medium effect that is moderated by variables such as domain subject, application type, intervention duration and learning environment.
Castillo-Manzano et al. (2016)	0.288 (k = 53)	14,963 adult learners in formal and informal settings across disciplines	Tests the effectiveness of audience response system (ARS) ("clickers")	An MA synthesising 53 ESs from 51 studies published between 2008 and 2015, addressing the impact of ARS on academic performance. Findings revealed a positive weak mean ES that is moderated by the context within which they are used, where non-university settings appeared to be more favourable for learning outcomes.
Hunsu et al. (2016)	0.050 (k = 86)	26,095 post-secondary students across subject matters	Tests the effectiveness of ARS ("clickers")	An MA synthesising 86 ESs from studies published between 2000 and 2014, addressing the use of ARS in education. Findings showed a small but significant ES resulted from using ARS-based technologies on some cognitive and non-cognitive learning outcomes. Moderator analysis revealed that knowledge domain, class size and the use of clicker questions are some factors that significantly moderated the ESs.
Sala et al. (2018)	0.070 (k = 359)	3,285 video-game players across grade levels and subject matters	Multipurpose use of DGBL (formal & informal settings)	An MA synthesising 359 achievement ESs from studies published between 1987 and 2016, addressing the effects of video game training on participants' cognitive ability. Findings revealed a null overall ES, indicating the absence of a relationship between overall cognitive ability and video game skill.

Steenbergen-Hu & Cooper (2014)	0.350 (k = 37)	Unspecified number of college students across disciplines	Tests effectiveness of multipurpose ITSs	An MA synthesising 37 achievement ESs from 35 studies published between 1990 and 2012, addressing the effectiveness of ITSs for college students. Findings revealed a moderate positive effect of ITSs on college students' academic learning. Moderator analyses revealed that an ITS was less effective than human tutoring but outperformed all other instruction methods and learning activities.
Cho et al. (2018)	0.501 (k = 22)	2,191 second language learners of all grade levels and age groups	Tests effectiveness of multipurpose MALL	An MA synthesising 22 achievement ESs from 20 studies published between 2005 and 2017, addressing the effectiveness of mobile technologies in language learning. Findings revealed a moderate positive overall effect. Further analyses revealed that test type and source of the study were the only significant moderators.
Merchant et al. (2014), sub-collection one	0.510 (k = 13)	3,081 students across grade levels and subject matters	VR: various educational games	An MA synthesising 13 achievement ESs from studies published between 1988 and 2011, addressing the effect of instructional design principles in the context of VR technology-based instruction. Results suggest games, simulations and virtual worlds were effective in improving learning outcome gains with a moderate positive mean ES. Moderator analysis indicated that games show higher learning gains than simulations and virtual worlds.
Merchant et al. (2014), sub-collection two	0.410 (k = 56)	5,351 students across grade levels and subject matters	VR: simulations for cognitive support for learning	An MA synthesising 56 achievement ESs from studies published between 1988 and 2011, addressing the effect of instructional design principles in the context of VR technology-based instruction. Results suggest games, simulations and virtual worlds were effective in improving learning outcome gains with a moderate positive mean ES. Moderator analysis indicated that games show higher learning gains than simulations and virtual worlds.
Sung et al. (2015)	0.531 (k = 223)	9,154 students of all grade levels in language arts (L1 & L2) courses	Tests effectiveness of multipurpose MALL	An MA synthesising 223 achievement ESs from 44 studies published between 1993 and 2013, addressing technology use with language learning. Findings revealed a moderate positive mean ES for mobile-device-assisted language instruction.
Sitzmann (2011)	0.277 (k = 39)	2,758 learners (post-secondary students and workplace trainees) across disciplines	Specific CAI applications: simulations and simulation games	An MA synthesising 39 achievement ESs from studies published between 1976 and 2009, addressing the instructional effectiveness of computer-based simulation games. Findings revealed a small positive ES, which was moderated by characteristics of the simulation games and the instructional context.
Wenting et al. (2014)	0.440 (k = 66)	11,296 students across grade levels and disciplines	Tests effectiveness of multipurpose ITSs	An MA synthesising 66 ESs from studies published between 1990 and 2012, addressing the use of ITS. Findings revealed that an ITS is associated with greater achievement in comparison with teacher-led, large-group instruction, non-ITS computer-based instruction, and textbooks or workbooks.

Schmid et al. (2014)	0.270 (*k* = 879)	58,585 post-secondary students across subject matters	Multipurpose various CAI applications	An MA synthesising 879 achievement ESs from studies published between 1990 and 2009, addressing technology use in post-secondary education. Findings revealed a small positive mean ES and moderator analyses revealing that cognitive support applications result in significantly higher effects.
Chien et al. (2016)	0.490 (*k* = 22)	3,963 students of all grade levels across subject matters	Tests the effectiveness of ARS ("clickers")	An MA synthesising 22 achievement ESs from studies published between 2003 and 2012, addressing whether clicker-integrated instruction is effective in enhancing students' learning gains. Findings revealed a positive mean ES with further analysis revealing that engaging students in explaining and justifying their answers to clicker questions is associated with more positive ESs.
Fengfeng et al. (2006)	0.354 (*k* = 22)	Unspecified number of students across grade levels and subject matters	Specific CAI applications: use of animation for content presentation	An MA synthesising 22 achievement ESs from 11 studies published between 1988 and 2004, addressing the effectiveness of animation in facilitating multi-level learning. Results indicated that animation generally has a small positive effect on facilitating multi-level learning. Results indicated that animation is not equally effective for different grade levels.
Moran et al. (2008)	0.489 (*k* = 89)	Unspecified number of middle school (Grades 6–8) students receiving reading instruction	CAI: multipurpose use of digital tools in literacy instruction	An MA synthesising 89 achievement ESs from 20 studies published between 1988 and 2005, addressing digital tools and learning environments to enhance literacy acquisition for middle school students. Findings demonstrate that technology can have a positive effect on reading comprehension, with little research focusing on other important aspects of reading, such as metacognitive, affective and dispositional outcomes.
Thompson & von Gillern (2020)	0.685 (*k* = 20)	1.298 second language learners of all grade levels and age groups	Specific CAI application: using DGBL in L2 vocabulary acquisition	An MA synthesising 20 achievement ESs from 19 studies published between 2006 and 2017, addressing the efficacy of DGBL in English as a second language vocabulary acquisition process. Findings revealed a moderately large effect.
Steenbergen-Hu & Cooper (2013)	0.090 (*k* = 34)	36,829 K–12 learners of mathematics	Tests effectiveness of multipurpose ITS	An MA synthesising 34 achievement ESs from 26 studies published between 1997 and 2010, investigating the effectiveness of ITSs on K–12 students' mathematical learning. Findings revealed a very small positive effect, with the impact being greater when the interventions lasted less than a school year.

Peng et al. (2021)	0.928 (k = 17)	Unspecified number of second language learners of all grade levels and age groups	Tests effectiveness of MALL	An MA synthesising 17 achievement ESs from 17 studies published between 2008 and 2018, addressing the effect of MALL. Findings revealed a large effect for mobile technologies in language learning, with type of activities, modality of delivery and duration of treatment moderating the ES.
Castro-Alonso et al. (2021)	0.200 (k = 32)	2,104 students across grade levels and subject matter	Specific CAI application: multipurpose use of "pedagogical agents"	An MA synthesising 32 ESs from 32 studies published between 2012 and 2019, addressing the effect of multimedia pedagogical agents on learning. Findings revealed a small overall effect, further indicating that 2D agents tended to be more effective than 3D agents.
Shi et al. (2021)	0.639 (k = 23)	954 students from pre-kindergarten to Grade 12 across study fields/ subject matters	Specific CAI application: multipurpose use of interactive whiteboards (IWB)	An MA synthesising 23 ESs from 23 studies published between 2010 and 2018, addressing the effectiveness of IWB-based instruction on cognitive learning outcomes. Findings revealed a positive effect, and moderator variable analysis indicated that the pedagogical approach and the year of publication significantly moderate the ES. Furthermore, IWB-based instruction was most effective when applied with an independent learning approach.
Mei-Mei (2019)	0.933 (k = 84)	6,296 second language learners of all grade levels	Multipurpose CAI for language learning	An MA synthesising 84 achievement ESs from 84 studies published between 1990 and 2015, addressing technology-enhanced language learning. Results revealed a large mean ES and indicated that the effects may be larger for small samples of participants recruited from higher education, with the instruction being delivered via general-purpose applications on mobile phones.
Hassan Saleh & Ahmed Abdulateef Al (2019)	0.666 (k = 20)	1,014 second language learning of all grade levels and age groups	Specific CAI applications for pronunciation training in L2 learning	An MA synthesising 20 achievement ESs from 20 studies published between 1992 and 2019, addressing the effectiveness of using a computer in pronunciation learning and training. Findings revealed a positive effect on L2 pronunciation, with computer use being equally effective for young and adult learners, but more effective for beginner and intermediate learners than advanced learners.
Chen, Tseng et al. (2018)	0.984 (k = 10)	635 second language learning of all grade levels and age groups	Specific CAI application: using DGBL in L2 vocabulary acquisition	An MA synthesising 10 achievement ESs from 10 studies published between 2006 and 2014, addressing the effects of DGBL on vocabulary. Results revealed a strong positive mean ES, indicating that the impacts vary with game design features but not with learners' age or linguistic background.

Kunkel (2015)	0.140 (k = 13)	17,037 students (from preschool to secondary school) receiving instruction in reading	Multipurpose CAI applications for reading instruction	An MA synthesising 13 achievement ESs from 13 studies published between 2000 and 2014, addressing the effectiveness of CAI reading outcomes of students in preschool through secondary school. Results indicated that the mean effects for students receiving reading CAI were small, positive and statistically significant when compared to control groups receiving no treatment or non-reading CAI.
Slavin & Lake (2008)	0.188 (k = 38)	Unspecified number of elementary school students studying mathematics	Multipurpose CAI applications for instruction in mathematics	An MA synthesising 38 achievement ESs from 38 studies published between 1990 and 2005, addressing the effects of CAI on elementary students' achievement in mathematics. The study also addressed the effects of mathematics curricula and instructional process programmes. Findings revealed moderate effects of CAI, with the strongest positive effects being for instructional process approaches such as forms of co-operative learning, classroom management and motivation programmes, and supplemental tutoring programmes.
Sung et al. (2016)	0.523 (k = 419)	18,749 students across grade levels and subject matters	Multipurpose use of various mobile devices	An MA synthesising 419 achievement ESs from 110 studies published between 1993 and 2013, addressing integrated mobile devices in teaching and learning. Results revealed a moderate mean ES, with higher ESs being found with inquiry-oriented learning applications, informal educational environments and medium- and short-duration interventions.

Notes: All effect sizes (ESs) are expressed in Hedges' *g* metric — as used in all analyses

AR: augmented reality
ARS: audience response system
CAI: computer-assisted instruction
CALL: computer-assisted language learning
CMC: computer-mediated communication
CSCL: computer-supported collaborative learning
CT: computer technology
DGBL: digital game-based learning
EFL: English as a foreign language
ICT: information and communications technology
ITS: intelligent tutoring system
IVR: immersive virtual reality
L1: native language
L2: second (foreign) language
MA: meta-analysis
MALL: mobile assisted language learning
MCSCL: mobile computer-supported collaborative learning
PDA: personal digital assistant
SCMC: synchronous computer-mediated communication
SLA: second language acquisition
STEM: science, technology, engineering, mathematics
VR: virtual reality

Table 6. Key descriptors of meta-analyses in the Online Learning collection (k = 11)

Study ID	Reported effect size (ES)	Sample & context characteristic	Technology type & functionality	Review focus & summary of findings
Bernard, Abrami, Wade et al. (2004)	0.013 (*k* = 318)	54,775 learners of all grade levels and subject matters	All forms (synchronous & asynchronous) of distance education (DE)	An MA synthesising 318 achievement ESs from 232 studies published between 1985 and 2002, addressing DE in comparison to face-to-face (F2F). Results indicated an average of approximately 0 with a wide variability. Further analysis revealed that synchronous applications favoured classroom instruction, while asynchronous applications favoured DE.
Ziegler (2016)	0.129 (*k* = 37)	623 second language learners of all ages	Synchronous computer-mediated communication (SCMC) (e.g., text chats, videoconferencing) to enable and support interactions among learners	An MA synthesising 37 ESs from 14 studies published between 1990 and 2012, addressing effectiveness of interaction in SCMC and F2F with second language learning. Results revealed no significant differences between SCMC and F2F, thus indicating that the mode of communication has no impact on the positive developmental benefits associated with interaction.
Williams (2006)	0.148 (*k* = 34)	2,784 participants of post-secondary and professional health education programmes	Tests effectiveness of several forms (by instructional design) of DE	An MA synthesising 34 ESs from 25 studies published between 1990 and 2005 and addressing effectiveness of DE in allied health professions. Findings revealed a very small positive effect for DE learners in comparison to F2F. Findings also revealed that in DE settings, working professional students outperformed graduate and undergraduate students. Open learning and synchronous instruction were the most effective DE models of instruction.
Means et al. (2013)	0.050 (*k* = 27)	Estimated over 2,200 K–12 students across subject matters	Different forms of online learning with various functions compared to F2F instruction	An MA synthesising 23 achievement ESs from studies published between 1996 and 2008, comparing learning outcomes for either fully online or blended learning conditions with those of F2F classroom instruction. Findings revealed that on average, students in online learning conditions performed modestly better than those receiving F2F instruction. The advantage over F2F classes was significant in those studies comparing blended learning with traditional F2F instruction but not in those studies contrasting purely online with F2F conditions.
Cavanaugh et al. (2004)	-0.028 (*k* = 116)	7,561 K–12 students across subject matters	Multifunctional Web-based instruction	An MA synthesising 116 achievement ESs from studies published between 1999 and 2004, addressing the effectiveness of Web-delivered K–12 DE programmes. The analysis revealed that DE may have the same effect on measures of student academic achievement when compared to traditional instruction. No factors related to either significant positive or negative effects were found.
Sitzmann et al. (2006)	0.149 (*k* = 71)	10,920 adult participants in post-secondary education and workplace training	Multifunctional Web-based instruction (WBI) compared to F2F instruction across subject matters	An MA synthesising 71 achievement ESs from studies published between 1996 and 2005, addressing the effectiveness of WBI in comparison to classroom instruction (CI). The overall results indicated WBI was more effective than CI for teaching declarative knowledge, and WBI and CI were equally effective for teaching declarative knowledge when the same instructional methods were used to deliver both WBI and CI. Finally, WBI was more effective than CI for teaching declarative knowledge when Web-based trainees were provided with control, in long courses, and when trainees practised the training material and received feedback during training.

Gegenfurtner & Ebner (2019)	0.131 (k = 14)	1,291 adult learners participating in webinars in higher and professional education	Multipurpose webinars across subject matters	An MA synthesising 14 ESs from studies published between 2010 and 2017, addressing effectiveness of webinars and videoconferences in promoting student knowledge and skills. Findings revealed a small mean ES, suggesting that webinars were slightly more effective than control online asynchronous learning management systems and offline F2F CI.
Allen et al. (2004)	0.238 (k = 38)	2,215 learners of unspecified grade level and presumably across subject matters	Various forms of DE, including online learning and videoconferencing	An MA synthesising 38 achievement ESs from studies published between 1983 and 2000, comparing the performance of students in DE vs traditional classes. The average effect demonstrates that DE course students slightly outperformed traditional students on exams and course grades. The average effect was heterogeneous, with analyses revealing no moderating variables.
Wandera (2017)	0.151 (k = 8)	Estimated over 3,500 predominantly post-secondary students across subject matters	Various forms of online learning compared to F2F instruction	An MA synthesising 8 achievement ESs from studies published between 2007 and 2017, comparing learning outcomes of online learning vs F2F education. Findings revealed a small positive mean ES in favour of online learning.
Roberts (2011)	0.777 (k = 59)	5,779 adult learners across subject matters	Web-based online learning only	An MA synthesising 59 achievement ESs from studies published between 1998 and 2010, addressing the efficacy of Web-based online learning among adult learners. A strong positive mean ES in favour of Web-based adult distance instruction was found. Results suggest Web-based DE appears to have improved over time and that independent study, behaviourist instructional strategies, instructor-moderated collaboration, provision of formative feedback and the use of multimedia are more effective practices to use in Web-based DE with adult learners.
Shachar & Neumann (2010)	0.256 (k = 125)	20,800 participants (predominantly college students and professionals)	Various forms of online learning	An MA synthesising 125 achievement ESs from 125 studies published between 1990 and 2009, addressing the differences between the academic performance of students enrolled in DE courses, relative to those enrolled in traditional settings. Findings indicated that students taking courses by DE outperformed their student counterparts in the traditionally instructed courses.

Notes: * Two studies from 2004 were not assigned the MMRQG “High” category only because of their use of the fixed effect analytical model (at that time this choice was quite common, though), otherwise meta-analyses by Bernard, Abrami, Lou et al. (2004) and Cavanaugh et al. (2004) were carried out with commendable methodological quality.

CI: classroom instruction
CMC: computer-mediated communication
DE: distance education
ES: effect size

F2F: face-to-face (in-class) instruction
MA: meta-analysis
SCMC: synchronous computer-mediated communication
WBI : Web-based instruction

Table 7. Key descriptors of meta-analyses in the Blended Learning collection (k = 13)

Study ID	Reported effect size (ES)	Sample & context characteristics	Technology type & functionality	Review focus & summary of findings
Bernard, Borokhovski, Schmid, Tamim & Abrami (2014)	0.334 (k = 117)	Post-secondary students across subject matters	Multipurpose blended learning (BL) (no more than 50% of time for the online component)	An MA synthesising 117 achievement ESs from studies published between 1990 and 2010, addressing the effectiveness of BL in higher education. Results indicate that, in terms of achievement outcomes, BL conditions exceed classroom instruction (CI) conditions by about one-third of a standard deviation and that the kind of computer support used and the presence of one or more interaction treatments are likely to enhance student achievement.
Means et al. (2013)	0.350 (k = 23)	Estimated over 3,300 K–12 students across subject matters	Different forms of BL with various functions compared to face-to-face (F2F) instruction	An MA synthesising 23 achievement ESs from studies published between 1996 and 2008, comparing learning outcomes for either fully online or BL conditions with those of F2F CI. Findings revealed that on average, students in online learning conditions performed modestly better than those receiving F2F instruction. The advantage over F2F classes was significant in those studies contrasting BL with traditional F2F instruction but not in those studies contrasting purely online with F2F conditions.
Li et al. (2019)	0.662 (k = 8)	574 medical school (post-secondary) students across subject matters	Various forms of BL with multiple purposes	An MA synthesising 8 achievement ESs from 8 studies published between 2007 and 2017, addressing the effects of BL on nursing students' knowledge, skills and satisfaction. Findings revealed that compared with traditional teaching, BL could effectively improve nursing students' knowledge.
Vo et al. (2017)	0.385 (k = 51)	7,033 post-secondary students across subject matters	Various forms of BL with multiple purposes	An MA synthesising 51 achievement ESs from studies published between 2000 and 2015, addressing the impact of BL on the academic achievement of post-secondary students. The results indicated that BL demonstrates a small summary effect compared to traditional teaching methods. A significantly higher mean ES was found in science, technology, engineering and mathematics (STEM) disciplines compared to that of non-STEM disciplines.
Sitzmann et al. (2006)	0.336 (k = 33)	6,799 adult participants in post-secondary education and workplace training	Web-based supplements to F2F instruction across subject matters	An MA synthesising 33 achievement ESs from studies published between 1996 and 2005, addressing the effectiveness of Web-based instruction (WBI) relative to CI. The overall results indicated WBI was more effective than CI for teaching declarative knowledge. Both delivery media were equally effective for teaching procedural knowledge. However, WBI and CI were equally effective for teaching declarative knowledge.

Wandera (2017)	0.468 (k = 21)	Estimated over 2,000 predominantly post-secondary students across subject matters	Various forms of multipurpose blended learning	An MA synthesising 21 achievement ESs from studies published between 2007 and 2017, comparing learning outcomes of BL vs F2F education. Findings demonstrated a moderate positive mean ES in favour of BL.
Låg & Sæle (2019)	0.350 (k = 272)	Unspecified number of students at all levels (but predominantly post-secondary) across subject matters	Tests the overall effectiveness of flipped classrooms (i.e., presentation format) compared to traditional lectures	An MA synthesising 272 achievement ESs from studies published between 2010 and 2017, comparing flipped classroom teaching with traditional, lecture-based teaching. Results revealed a small mean ES in favour of the flipped classroom. There is some support for the notion that the positive impact on learning may increase slightly if testing student preparation is part of the implementation.
Cheng et al. (2019)	0.193 (k = 115)	7,912 students across grade levels and subject matters	Tests the overall effectiveness of flipped classrooms (presentation format)	An MA synthesising 115 achievement ESs from 55 studies published between 2000 and 2016, addressing the effect of the flipped classroom instructional strategy on student learning outcomes. Findings revealed a statistically significant small mean ES in favour of the flipped classroom instructional strategy. The ESs were significantly moderated by subject area, such as mathematics, science, social sciences, engineering, arts and humanities, health and business.
Talan & Batdi (2020)	0.462 (k = 71)	Unspecified number of students across grade levels and subject matters	Tests the overall effectiveness of flipped classrooms (presentation format)	An MA synthesising 71 achievement ESs from 64 studies published between 2013 and 2019, addressing the effectiveness of the flipped classroom model (FCM) in an educational setting. The results of the study indicated that FCM had a positive impact on academic success in general.
Zheng, Bhagat et al. (2020)	0.663 (k = 95)	15,386 students across grade levels and subject matters	Tests the overall effectiveness of flipped classrooms (presentation format)	An MA synthesising 95 achievement ESs from 95 studies published between 2013 and 2019, addressing the overall effectiveness of the flipped classroom on students' learning achievement and motivation. Results revealed that the flipped classroom approach had a moderate ES for learning achievement. Findings indicated that sample size, intervention durations and sample regions significantly moderated the effect sizes.

Lo & Hew (2019)	0.289 (k = 29)	5,329 engineering (high school and post-secondary) students	Tests the overall effectiveness of flipped classrooms (presentation format)	An MA synthesising 29 achievement ESs from 29 studies published between 2008 and 2017, addressing the effects of flipped classrooms on student achievement in engineering education. Findings revealed an overall significant effect in favour of the flipped classroom over traditional lectures. Moderator analyses indicated that the effect of the flipped classroom was further enhanced when instructors offered a brief review at the beginning of CI.
Shahnama et al. (2021)	1.240 (k = 69)	3,338 second language learners of all age groups	Tests the effect of flipped classrooms (presentation format) on various language skills (reading, writing, etc.)	An MA synthesising 53 achievement ESs from 27 studies published between 2011 and 2019, addressing the effectiveness of a flipped learning approach in improving students' achievements. Findings indicated that the influence of flipped learning on students' achievements was large and positive. Moderator variable analysis indicated that the flipped group performed better when they received material ahead of time to help them prepare for activities and exercises.
Zhu (2021)	0.536 (k = 53)	Unspecified number of K–12 students across subject matters	Tests the overall effectiveness of flipped classrooms (presentation format)	An MA synthesising 53 achievement ESs from 27 studies published between 2007 and 2017, addressing the effect of flipped instruction on K–12 students' academic achievement. Findings revealed a positive moderate strength mean ES, indicating that flipped instruction promoted students' academic achievement moderately better than traditional CI. It was found that only the variable publication type explains a significant degree of ES heterogeneity.

Notes:
BL: blended learning
CI: classroom instruction
CMC: computer-mediated communication
ES: effect size
F2F: face-to-face
FCM: flipped classroom model
MA: meta-analysis
STEM: science, technology, engineering and mathematics
WBI: Web-based instruction

The Overall Effect Sizes

Table 8 (below) summarises the overall findings (students' learning achievement outcomes) for the three types of instructional settings:

1. In-class technology integration (k = 108, based on over a million learners; the numbers of participating students are estimated, as samples could partly overlap across included meta-analyses)
2. Online learning (k = 12, based on over 119,000 learners)
3. Blended learning, including the use of flipped classrooms (k = 14, based on over 53,000 learners)

The corresponding weighted average effect sizes, depicted here as g++ (to underline that they represent a synthesis of summaries in this second-order meta-analysis in contrast to g+ as the aggregated effects in the included first-order meta-analyses), are reported according to both random effects and fixed effect analytical models. Given the apparent non-uniformity of studies in our collection, the former provides more accurate estimates of the effects in the studied populations, while the latter serves primarily to estimate the heterogeneity of the respective effect size distributions. These statistics are reported in the last line of each subsection of the table.

Table 8. Overall weighted average effect size for achievement outcomes and associated heterogeneity statistics: Preliminary data set

Population estimates	*k*	*g*++	SE	Lower 95th confidence interval	Upper 95th confidence interval
In-class technology integration					
Random effects model	108	0.484**	0.03	0.43	0.54
Fixed effect model	108	0.450**	0.02	0.42	0.48
Heterogeneity analysis	Q_T = 253.42 (df = 107), p < .001, I^2 = 57.78				
Online learning					
Random effects model	12	0.232**	0.09	0.06	0.40
Fixed effect mode	12	0.156*	0.03	0.06	0.25
Heterogeneity analysis	Q_T = 27.25 (df = 11), p < .01, I^2 = 59.63				
Blended learning					
Random effects model	14	0.478**	0.06	0.33	0.63
Fixed effect model	14	0.442**	0.02	0.35	0.53
Heterogeneity analysis	Q_T = 28.05 (df = 13), p < .01, I^2 = 53.66				

* p < .05; ** p < 0.01

“International”’ vs “Single Region” Meta-analyses

Before proceeding further with the analyses of each collection, we decided to explore one specific characteristic of the included meta-analyses that could influence the entire data set: specifically, if the meta-analyses were based on data gathered internationally as opposed to primary research collected exclusively from a single region. This characteristic seemed to be of importance as there were 16 such meta-analyses in total, across the three categories of instructional settings.

As mentioned in the Methodology section, at least two major considerations must be taken into account when determining whether to use the “single region” category of meta-analyses. First, the primary studies they summarise are likely to be located and reviewed by larger “international” meta-analyses, and excluding them would partly address the issue of accumulating dependency (i.e., accounting for the same primary studies more than once). Second, our preliminary review of the collected data shows that these exclusively “single region” meta-analyses tend to include a large portion of master’s and PhD theses from local higher education institutions (those are more dependent on regional educational contexts, typically based on smaller samples, implemented with local issues in focus, etc.). As such, these region-based meta-analyses could be more susceptible to a potential bias, which may result in substantial overestimation of the corresponding effect sizes (see, for example, Bernard et al., 2018; Pigott & Polanin, 2020; Tamim et al, 2021).

To test this possibility, we ran a moderator variable analysis (according to the mixed effects analytical model) on the entire data set of 134 effects, comparing meta-analyses conducted “internationally” and by “single region.” The results are summarised in Table 9 (below).

Table 9. Moderator variable analysis (mixed effects model): Meta-analyses conducted internationally and by single region

Levels of moderator variables	Number of effects (k)	Category average ES (g++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{Between}$
Intervention type: Content-specific vs developmental					
International data	118	0.398***	0.35	0.45	
Single region data	16	0.852***	0.73	0.97	
Between groups: $df = 1$					47.68, $p < .01$

*** $p < .001$

The data reported in Table 9 confirm our assumption about the likely incompatibility of the "international" and "single region" meta-analyses: the weighted average effect size for the latter (g++ = 0.852) is more than twice that of the former (g++ = 0.398), and this difference is statistically significant ($p < .01$). Subsequently, we decided to exclude "single region" studies from further consideration. There were 14 of these meta-analyses in the In-Class Technology Integration collection, one in the Online Learning collection and one in the Blended Learning collection. We subsequently adjusted the corresponding weighted average effect sizes, as presented in Table 10 (random effects model only).

Table 10. Overall weighted average effect size for achievement outcomes: Final data set

Population estimates	*k*	*g*++	SE	Lower 95th confidence interval	Upper 95th confidence interval
In-class technology integration					
Random effects model	94	0.415**	0.03	0.36	0.47
Online learning					
Random effects model	11	0.169*	0.07	0.03	0.31
Blended learning					
Random effects model	13	0.470**	0.08	0.32	0.62

* $p < .05$; ** $p < 0.01$

The adjusted effect sizes (notably smaller than those presented in Table 8 prior to the removal of 16 "single region" meta-analyses) were used in all subsequent data analyses (including publication bias, sensitivity and all classes of available moderator variable analyses), organised and reported by collection with the reduced number of cases in each: In-Class Technology Integration (k = 94), Online Learning (k = 11) and Blended Learning (k = 13).

In-Class Technology Integration Collection

We first carried out a complete set of publication bias analyses. Their outcomes are presented below.

Publication bias analyses

1. DUVAL AND TWEEDIE'S TRIM AND FILL ANALYSIS

Duval and Tweedie's trim and fill analysis and the related funnel plot are shown in Figure 2. Under the random effects model, the point estimate and 95% confidence interval (in parentheses) for the aggregated data is g++ = 0.415 (CI_{L-U} = 0.36, 0.47). Using these parameters while examining the funnel plot, it is suggested that some studies could be missing from the negative (< 0.00) or the left side of the distribution.

To balance the distribution, the trim and fill method statistically imputes these potentially "missing" studies (k = 17). The resulting adjusted point estimate is g++ = 0.347 (CI_{L-U} = 0.29, 0.40), a decrease of 0.07 from the calculated empirical value.

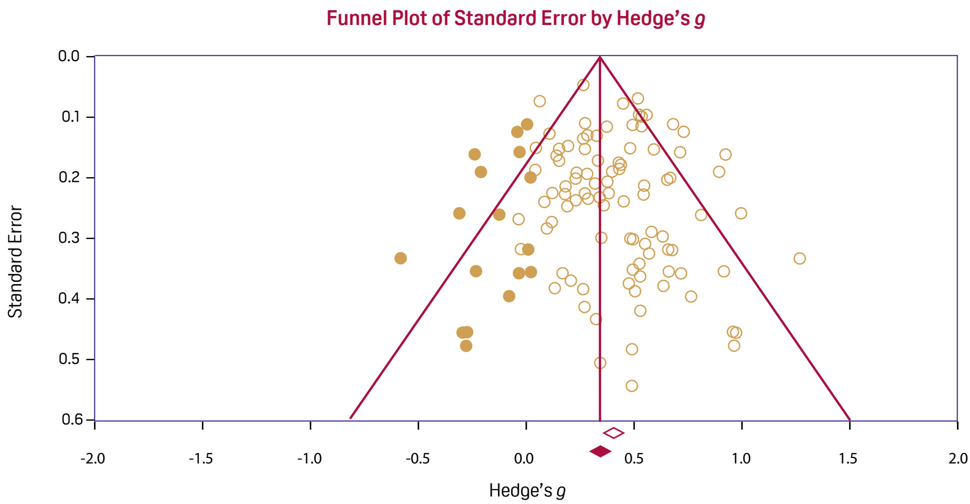

Figure 2. Funnel plot of achievement effects sizes for the In-Class Technology Integration collection of meta-analyses.

Note: White circles indicate existing empirical studies (k = 94) and coloured circles are imputed studies that are presumed missing (k = 17).

There are at least two possible explanations for the difference between the empirical mean (observed prior to adjustment) and the mean projected through the trim and fill analytical routine (adjusted through imputation to balance the distribution). One, the empirical average could be correct, meaning that there truly is a preponderance of positive studies compared with negative studies in the literature, and as such, the imputation of "missing" studies is just a statistical precaution. Two, the research in which "missing" negative value studies were reported exists but was either not available from easily attainable sources (e.g., found only in conference proceedings or technical reports) or not made publicly available (i.e., not published or otherwise shared with the research community). The latter explanation constitutes the classic case of publication bias (also known as the file-drawer effect), when negative or non-significant findings are set aside or put on hold for further examination/replication by publishers and researchers alike. At this point it is not clear which explanation is correct.

Both values of the overall effect size are not only statistically significant but also quite robust, as confirmed by other analyses. However, in interpreting the findings, we would rely on more conservative, adjusted estimates.

2. CLASSIC FAIL-SAFE ANALYSIS

Classic fail-safe analysis revealed that 10,377 potentially "missing" null effect studies would be needed to render the observed overall weighted average effect size non-significant (i.e., to bring the calculated *p*-value above the pre-set alpha of 0.05). In addition, Orwin fail-safe analysis (Orwin, 1983) resulted in n = 280 — that is, the number of "missing" null effects that, if added to the existing collection, would bring the observed combined Hedges' g++ down to 0.1. In other words, despite some degree of publication bias, the aggregated effect of In-Class Technology Integration on students' learning achievement is considered to be moderately (in Cohen's terms, 1988) positive, statistically significant and fairly stable.

3. ONE-STUDY-REMOVED SENSITIVITY ANALYSIS

Finally, one-study-removed sensitivity analysis identified no potential outliers ("large leverage" studies that could disproportionally influence the overall outcomes). Subsequently, no further adjustment to the collection of k = 94 was made.

Table 11 (below) presents unadjusted (both random and fixed) and adjusted overall mean effect sizes (learning achievements) for the In-Class Technology Integration collection of meta-analyses.

Table 11. In-class technology integration overall weighted average effect size for learning achievement outcomes and associated heterogeneity statistics: Final data set

Population estimates	*k*	*g*++	SE	Lower 95th confidence interval	Upper 95th confidence interval
In-class technology integration					
Random effect model	94	0.415**	0.03	0.36	0.47
Fixed effect model	94	0.397**	0.02	0.36	0.43
Adjusted effect (trim and fill imputation)	111 (94+17)	0.347**		0.29	0.40
Heterogeneity analysis Q_T = 156.25 (df = 93), p < .001, I^2 = 40.48					

** p < 0.01

As Table 11 indicates, the distribution of the effects sizes was significantly heterogeneous, which warranted a further exploration of potential sources of systematic variation by means of moderator variable analyses.

Analysis of methodological moderator variables

First, we checked the specifics of the implementation of included primary meta-analyses in our collection. As described in the Methodology section, two major aspects of implementation were coded:

- Completeness of procedures for assessing the methodological quality of primary empirical studies within individual meta-analyses (e.g., coding and analysing research design, addressing the dependency issue) and precision/accuracy of their own implementation (e.g., conducting publication bias and sensitivity analyses).
- Whether sufficient rationale for the moderator variable analysis was provided and how many meaningful moderators were coded and explored. In the Methodology section we named complete absence of moderator variable analysis as one of the exclusion criteria. However, when moderator variables were analysed, we needed to see how justified and detailed this examination was in any given meta-analysis.

Table 12 (below) summarises the outcomes of the analyses of these methodological moderator variables. Individual Hedges' g++ reflects the magnitude of the effect associated with each level of the considered moderator variables, whereas the $Q_{Between}$ value is tested to verify whether differences between/among levels of these moderators are statistically significant.

Table 12. Moderator variable analysis (mixed effects model): Methodological characteristics, set (1)

Levels of moderator variables	Number of effects (k)	Category average ES (g++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{Between}$
Attention to methodological quality of included primary studies and own analytical procedures					
Complex (multiple) probes for methodological quality	34	0.302**	0.23	0.38	
Limited verification of methodological quality	41	0.525**	0.46	0.59	
No verification of methodological quality	19	0.406**	0.30	0.51	
Between groups: $df = 2$					19.85, $p < .01$
Degree of implementing moderator variable analysis					
Comprehensive moderator variable analyses	77	0.403**	0.35	0.46	
Limited moderator variable analyses	17	0.507**	0.38	0.63	
Between groups: $df = 1$					2.23, $p = .136$

** $p < .01$

Only analysis of the first moderator variable produced significant results with respect to the difference among its levels: $Q_{Between} = 19.85$ ($p < .01$). The source of this difference is primarily associated with the fact that the meta-analyses with multiple checks for methodological quality produced aggregated effects sizes lower than those of meta-analyses with either limited or no such checks. However, the subsequent post-hoc pair-wise comparison was not statistically significant. These findings reflect standard observations in the methodological literature on meta-analysis (see, for example, Cheung & Slavin, 2016; Tamim et al., 2021), that methodologically more rigorous meta-analyses tend to produce less pronounced effects, whereas shortcomings in the methodology of study selection and analysis implementation are often associated with likely overinflated effect sizes. Comprehensiveness of the moderator variable analysis did not significantly affect the magnitude of the aggregated effect size: $Q_{Between}$ = 2.23 (p = .136).

Achievement outcome measure source/category

Somewhat related to the above was our analysis of the source/category of the achievement outcome measures specified in different meta-analyses. As mentioned in the Methodology section, we distinguished between those that use multiple/generic learning achievement measures and those that explicitly stated either the specific outcome type (e.g., declarative knowledge or reading comprehension) or the corresponding measure source (e.g., standardised tests). Though this information cannot be definitively classified as a methodological moderator variable, it is indirectly indicative of how focused a given meta-analysis was and how much attention it paid to methodology-related details of the included primary empirical studies. The results of this moderator variable analysis are presented in Table 13 and demonstrate the same tendency described in the subsection above. Higher effect sizes are typically associated with more generic information (in this case, the aggregated effect size for the outcome measures reported indiscriminately was g++ = 0.502), whereas low effect sizes are more likely to be based on more specific data that are described in greater detail (here the aggregated effect size for specified outcome measure source/category was g++ = 0.379). The difference between the two levels of this moderator variable was marginally significant: $Q_{Between}$ = 4.02, p = .045.

Table 13. Moderator variable analysis (mixed effects model): Outcome measure source/category, set (2)

Levels of moderator variables	Number of effects (*k*)	Category average ES (*g*++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{Between}$
Achievement outcome measure source/category					
Measure source/category specified	37	0.502**	0.39	0.61	
Achievement outcomes reported indiscriminately	57	0.379**	0.33	0.43	
Between groups: *df* = 1					4.02, *p* = .045

** *p* < .01

Other substantive and contextual moderator variables

As mentioned in the Methodology section and elsewhere (e.g., Tamim et al., 2011), any second-order meta-analysis is limited in its ability to explore and explain moderator variables by the information available in the included first-order meta-analyses — and more specifically, to what extent this information appears across these meta-analyses. Very few meta-analyses in the field specify, for example, the gender composition of the participating samples or pedagogical frameworks employed in the instructional practices. Conversely, common moderators such as grade level or subject matter often appear in educational meta-analyses. Typically, they are analysed as moderators, but using them across first-order meta-analyses is only possible when there is a sufficient number of meta-analyses specifically dedicated to a particular level of such moderators. With this challenge in mind, whenever possible we coded and then analysed the following moderator variables:

- Grade level: to reflect the learners' academic level and age
- Subject matter: to reflect the discipline or field of knowledge to be acquired
- Publication date and coverage: to reflect the range of periods during which the primary studies included in each meta-analysis were conducted and show the age of the studies
- Major function of technology used: to reflect the predominant objective(s)/ expectations of employing a particular type of educational technology

The results of this analysis are summarised in Table 14 (below) and then discussed. When necessary, additional follow-up analyses were carried out with some modifications in coding of specific categories of moderator variables to clarify their comparative effects on the corresponding point estimates. The outcomes of these follow-up analyses are summarised in Tables 14a–14c (pages 60–64).

Table 14. Moderator variable analysis (mixed effects model): Substantive and contextual study features, set (3)

Levels of moderator variables	Number of effects (k)	Category average ES (g++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{Between}$
Grade level					
All levels (ages)	42	0.478**	0.40	0.56	
Early grades (only pre-K/K/ elementary school)	4	0.361**	0.17	0.55	
All grades of compulsory School (K-12)	25	0.362**	0.26	0.47	
Above compulsory school (post-secondary & professional training)	16	0.347**	0.24	0.45	
Between groups: *df* = 3					5.28, p = .153
Subject matter: STEM vs non-STEM					
Across disciplines	35	0.382**	0.31	0.46	
STEM	26	0.360**	0.29	0.44	
Non-STEM	33	0.526**	0.42	0.63	
Between groups: *df* = 1					7.14, p = .028
Years of coverage: Last year of included studies					
Up to 2010 (inclusive)	18	0.304**	0.24	0.37	
2011 – onward	75	0.452**	0.39	0.51	
Between groups: *df* = 1					11.01, p = .001
Major function of technology employed					
Cognitive support: deep learning	13	0.451**	0.35	0.55	
Cognitive support: distributed cognition	2	0.330	-0.05	0.71	
Gaming	11	0.304**	0.12	0.49	
Interaction/collaboration	4	0.397**	0.15	0.65	
Presentation/access to information	7	0.354**	0.22	0.49	
Several (non-specified) functions	48	0.424**	0.38	0.47	
Several (specified) functions	9	0.488**	0.26	0.72	
Between groups: *df* = 6					3.47, p = .748

* p < .05; ** p < .01
Note: One case of N/A (data not available) was removed from each – Grade Level and Years of Coverage moderator analysis (i.e., summative k = 93).

Only two sets of the results showed statistically significant differences among the levels of the respective moderator variables, namely:

1. Non-STEM subject matter seemed to benefit from educational technology use more (g++ = 0.526, k = 33) than either STEM disciplines (g++ = 0.360, k = 26) or all-encompassing (Across Disciplines) educational technology use (g++ = 0.382, k= 35) — both corresponding post-hoc pair-wise comparisons were statistically significant (p < .05).
2. Meta-analyses that aggregated data from primary empirical studies published up to and including 2010 produced a significantly (p = .001) lower overall weighted average effect size (g++ = 0.304, k = 18) than the meta-analyses whose inclusion end-date extended beyond 2010 (2011 onward) (g++ = 0.452, k = 75). With respect to the former, we attempted to locate a more specific source of the difference by exploring particular disciplines or fields of knowledge. The results of this attempt are found in Table 14a.

The significant difference relating to time of publication of included primary empirical studies also deserves more detailed examination, as the end-date by itself does not fully represent the time frame of the respective research coverage. For example, some meta-analyses completed after 2010 covered several decades of research, whereas others summarised only studies conducted within a decade prior to the meta-analyses' publication dates. As a result, we may want to further classify the meta-analyses in our collection to account for the scope of coverage and the overlap in coverage arising from the differences in scope. The outcomes of our follow-up analyses are shown in Table 14b.

Our analysis of the grade level moderator variable produced non-significant findings. The effect sizes were quite compatible across levels. Further exploration of this moderator variable was not possible based on the reported overall effects in the included meta-analyses. The only way to look closer at the effect of grade level would be to selectively summarise effects for this variable when reported as a subcategory in some of the meta-analyses. This merits further examination as a follow-up project but was beyond the scope of this current second-order meta-analysis.

Finally, though no statistically significant differences were detected among levels of the major function of technology use ($Q_{Between}$ = 3.47, p = .748), this moderator variable deserves more attention as it is probably the most substantive in this second-order meta-analysis, and it is instrumental for better understanding "what works" when technology is employed to improve/facilitate learning. With this in mind, we tried to simplify the coding of various categories of this moderator variable by combining the options "several non-specified functions" with the option "several specified functions," as these two showed no statistically significant difference in the associated effect sizes (g++ = 0.424, k = 48 and g++ = 0.488, k = 9, respectively) in a post-hoc pair-wise comparison. We also revisited two meta-analyses in which the main objective for technology use was coded as "cognitive support: distributed cognition." The meta-analysis by Sosa et al. (2011) emphasised the role of specialised computer applications (e.g., SPSS) for learning statistics, and the meta-analysis by Torgerson and Elbourne (2002) focused on using computers' word processing spell-checking capacity in composition writing, confirming our original coding which would prevent adding these two meta-analyses to any other category. Instead, they were excluded from our follow-up analysis of this moderator variable. Unfortunately, more refined coding of larger categories would not provide sufficient statistical power for analysis but would result in more categories with fewer cases per category. Instead, as reflected in the Methodology section, several functions, represented by just one or two cases (e.g., "access to information"), were already collapsed together with the more comprehensive "presentation" category. The results of analysing these refined categories of the major function(s) of technology use moderator variable are shown in Table 14c.

Results of a follow-up moderator variable analysis

First, we considered more refined coding of the subject matter, isolating specific components of both STEM and non-STEM categories as presented in the upper half of Table 8a. The difference among these modified levels was not significant ($Q_{Between} = 8.96$, $p = .111$). However, it is difficult to ignore the fact that the aggregated effect size for the all-inclusive "language arts" category of $g{+}{+} = 0.526$ ($k = 33$) was notably — although not significantly — higher than the effect sizes for any other category. Accordingly, we decided to further explore the source of variability within this category by analysing separately second language (L2) learning, first language (L1) competencies, and reading, especially early literacy. Also, the disciplines "medical education," "science" and "STEM" (as specified in the corresponding meta-analyses themselves) were represented by a very small number of cases ($k = 2$, $k = 3$, $k = 4$, respectively). Typically, as in most of our previous meta-analyses, we considered such numbers (under 5) insufficient for any meaningful interpretation, and hence these categories were removed from the subsequent analysis (see the lower half of Table 14a). This last round of analyses returned a highly significant level of heterogeneity across different categories: $Q_{Between} = 28.75$, $p < .001$. The series of post-hoc pair-wise comparisons indicated that the main source of this heterogeneity lay in the difference between "L2 learning" and all other categories (with the exception of "L1 learning" — the difference in effect sizes between these two was not statistically significant). Thus, it appears that language learning, and especially L2, benefits the most from integrating computer technologies in the educational process.

Table 14a. Moderator variable analysis (mixed effects model): Subject matter revisited

Levels of moderator variables	Number of effects (k)	Category average ES (g++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{between}$
Subject matter: Specific fields					
Across disciplines	35	0.382**	0.31	0.46	
Language arts (includes L1, L2 & reading)	33	0.526**	0.42	0.63	
Maths and statistics	17	0.319**	0.20	0.43	
Medical education	2	0.393*	0.02	0.77	
Science	3	0.311*	0.07	0.55	
STEM (as specified in actual meta-analyses)	4	0.345**	0.32	0.59	
Between groups: df = 5					8.96, p = .111
Subject matter: Modified categories					
Across disciplines	35	0.382**	0.31	0.46	
Second language (L2) learning	17	0.699**	0.57	0.73	
First language (L1) competencies	8	0.572**	0.43	0.71	
Reading	8	0.269	0.11	0.43	
Maths and statistics	17	0.319**	0.20	0.43	
Between groups: df = 4					28.75, p < .001

* p < .05; ** p < .01

Note: As several categories were excluded from the analysis, the total case count for the last part of the table is k = 85.

Next, we took a closer look at how the included meta-analyses covered research conducted at different points in time. As stated earlier, the overlap in included primary studies was taken into consideration in this follow-up analysis, resulting in the creation of the following three categories:

- **Dated meta-analyses:** meta-analyses that were published prior to and during 2010
- **Overarching meta-analyses:** meta-analyses of studies conducted over a wide range of years with a time overlap in the included studies
- **Recent meta-analyses:** meta-analyses published through the last decade and whose included primary studies were conducted within the same period

The outcomes of this refined analysis (Table 14b, upper half) indicate statistically significant differences in effect sizes among these three categories: $Q_{Between}$ = 11.44, p =.003. Though the absolute value of the effect size is the highest for the most recent meta-analyses (g++ = 0.467, k = 18), the true source of the difference is the overarching meta-analyses as the post-hoc pair-wise test shows, where its effect size (g++ = 0.450, k = 57) is significantly different ($p < .05$) from the effects size for "dated meta-analyses" while the larger in magnitude effect size for "recent meta-analyses" is not.

We were interested in understanding whether the size of the overlap in time makes any difference to the corresponding effect sizes. The lower half of Table 14b features the following four categories designed to address this issue:

- **Dated meta-analyses** with extensive coverage that include primary studies originating over more than 10 years
- **Recent meta-analyses** where the time coverage is under 10 years
- **Overarching meta-analyses** with extensive overlap in included studies with over 5 years of overlap across the cut-point of 2010/2011
- **Meta-analyses with limited overlap** in included studies (with 5 years of overlap across the cut-point of 2010/2011)

A few meta-analyses did not fall under any of the categories and were not included in this follow-up moderator variable analysis. The overall heterogeneity (difference in effect sizes among the explored categories) was statistically significant: $Q_{Between}$ = 13.67 (p = .003). The subsequent post-hoc pair-wise test indicated that only one of the comparisons was statistically significant ($p < .05$): the effect size of g++ = 0.559 (k = 18) for the "meta-analyses with limited overlap" (i.e., published in and after 2011 and going back prior to 2010 by 5 years or less) was significantly different from the effect size of g++ = 0.301 (k = 17) for "dated meta-analyses with extensive coverage."

Table 14b. Moderator variable analysis (mixed effects model): Publication date and research coverage revisited

Levels of moderator variables	Number of effects (k)	Category average ES (g++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{Between}$
Years of coverage: With and without the overlap in time					
Up to 2010 (inclusive), i.e. dated meta-analyses	18	0.304**	0.24	0.37	
2011 – Onward with some overlap in included studies, i.e., published prior to 2011	57	0.450**	0.38	0.52	
2011 – Onward only, i.e., recent meta-analyses	18	0.467**	0.34	0.59	
Between groups: df = 2					11.44, p = .003
Scope of coverage: Extensive vs limited					
Dated meta-analyses with extensive coverage	17	0.301**	0.24	0.36	
Recent meta-analyses (coverage under 10 years)	17	0.387**	0.25	0.52	
Meta-analyses with extensive overlap	39	0.415**	0.34	0.49	
Meta-analyses with limited overlap	18	0.559**	0.43	0.69	
Between groups: df = 3					13.67, p = .003

* p < .05; ** p < .01
Note: Removal of one case of N/A (data not available) from this analysis resulted in the total number of cases k = 93 in the upper half of the table. The lower half of the table contains only k = 91 cases, as several categories represented by one or two cases were excluded.

In addition, we supplemented these findings by meta-regression analyses (method of moments analytical model) of two continuous moderator variables: publication date (of each included meta-analysis itself) and publication coverage (the range of publication dates of studies included in the respective meta-analysis) as predictors and with the weighted average effect size (g++) as the criterion variable (Figures 3 and 4, respectively). The first regression line with a small positive slope of 0.014 was marginally significant (p = .012), indicating that with time, meta-analyses tend to produce effect sizes higher in magnitude. The second regression line is nearly symmetrically reversed — with the negative slope of -0.006 (p = .012). Taken all together, with the outcomes of the relevant categorical moderator variable analyses, these findings could indicate more recent research on technology integration in education's effectiveness for learning.

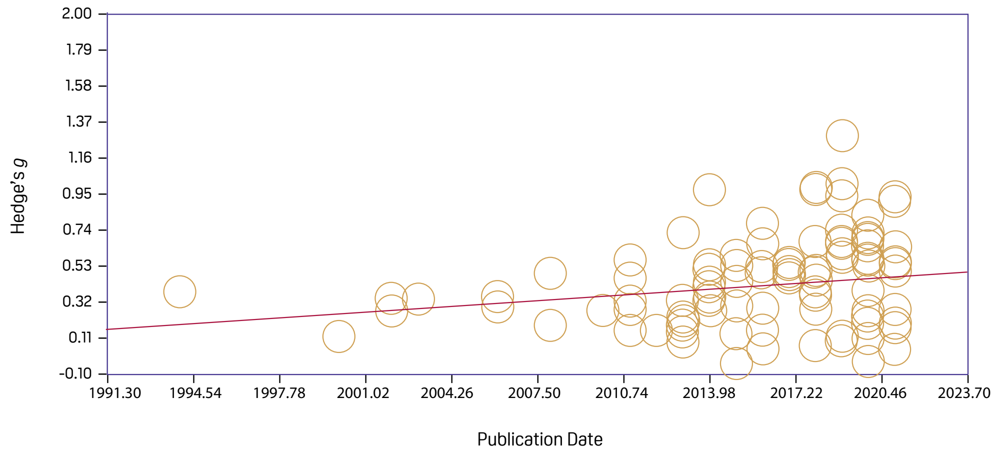

Figure 3. Meta-regression analysis of publication date.

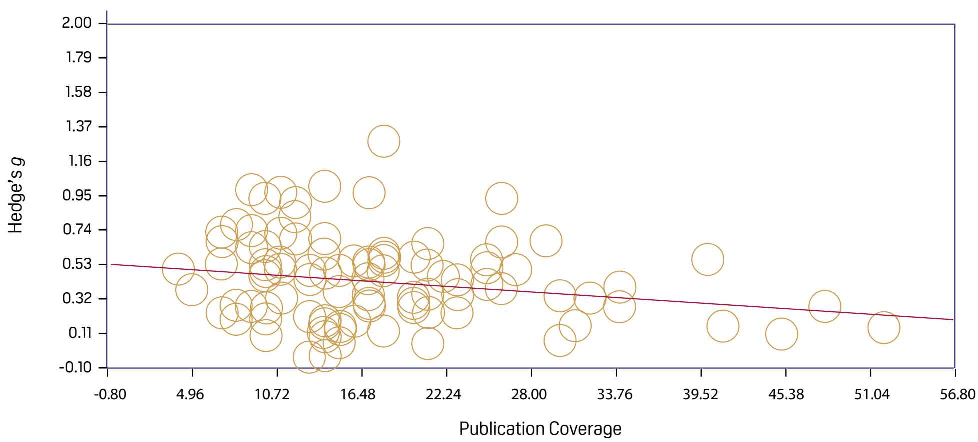

Figure 4. Meta-regression analysis of publication coverage.

Finally, the analysis of the major function of technology use moderator variable was revisited (Table 14c). Its levels were reconfigured, as described earlier, to produce the following categories: cognitive support (used primarily to facilitate understanding and promote deep learning, e.g., simulations), gaming (of various natures), interaction/collaboration (e.g., computer-mediated communication), presentation (understood broadly to encompass related functions, e.g., access to information) and a broad category of several functions (multiple objectives, both specified and not specified in the corresponding individual meta-analyses). This modification, however, failed to contribute to the explanation of variability in the aggregated effect sizes ($Q_{Between}$ = 2.97, p = .563). The only thing that is probably worth noting is that despite the overall non-significance of the findings, the effect of gaming on learning achievements (g++ = 0.304) is visibly lower than any other effect size in the collection, especially in comparison with the effect of cognitive support for deep learning (g++ = 0.451).

Table 14c. Moderator variable analysis (mixed effects model): Major function of technology revisited

Levels of moderator variables	Number of effects (k)	Category average ES (g++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{Between}$
Major function of technology employed					
Cognitive support: deep learning	13	0.451**	0.35	0.55	
Gaming	11	0.304**	0.12	0.49	
Interaction/collaboration	4	0.397**	0.15	0.65	
Presentation	7	0.354**	0.22	0.49	
Several functions	57	0.436**	0.37	0.50	
Between groups: df = 4					2.97, p = .563

** p < .01

Note: Reconfiguration of the categories resulted in the total number of cases k = 92

Online Learning Collection

As with the Technology Integration collection, our first set of analyses in the Online Learning collection included checks for publication bias and potential outliers. The outcomes of these analyses are presented below.

Publication bias and potential outliers analyses

1. DUVAL AND TWEEDIE'S TRIM AND FILL ANALYSIS

Duval and Tweedie's trim and fill analysis and the related funnel plot are shown in Figure 5. According to the random effects model, the observed point estimate and the corresponding 95% confidence interval (in parentheses) is g++ = 0.169 ($CI_{L\text{-}U}$ = 0.03, 0.31). The trim and fill analytical routine and examination of the funnel plot found no imbalance in the distribution of the effect sizes. The distribution was fairly symmetrical and in no need of imputation of any "missing" studies on its left side. In other words, no publication bias was detected.

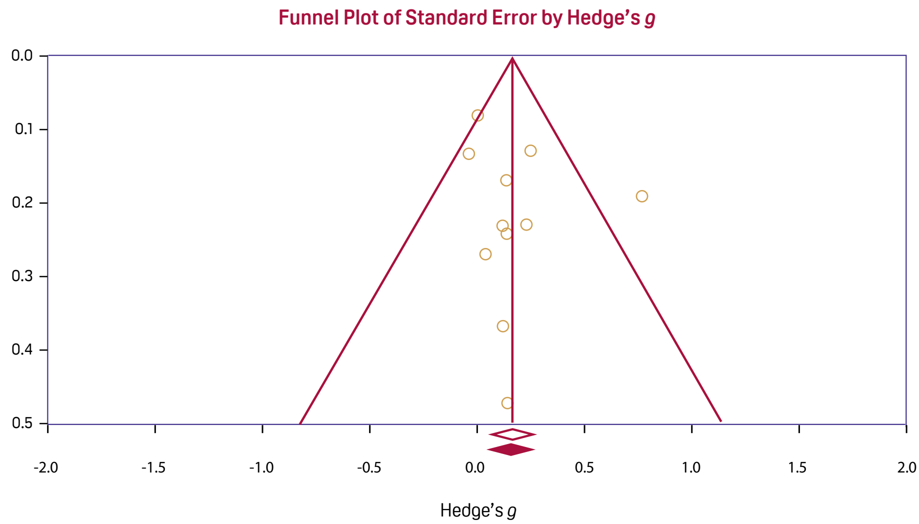

Figure 5. Funnel plot of achievement effects sizes for the Online Learning collection of meta-analyses

One effect size on the right side of the funnel plot is visibly dissociated from the rest of the distribution. So, even though its influence was not statistically significant (according to the trim and fill analyses), we assessed it as potentially meriting special attention.

Later in this section we present a short summary of the meta-analyses that reported this atypically high (for the Online Learning research) effect size of g++ = 0.777 (Roberts, 2011). Also, when it comes to summarising the data in the Online Learning collection of meta-analyses, we report the overall weighted average effect size both with and without this study, as reflected in Table 15.

2. CLASSIC FAIL-SAFE ANALYSIS

Classic fail-safe analysis indicated that 16 potentially "missing" null effect studies would be required to render the observed overall weighted average effect size non-significant (i.e., to bring the calculated p-value above the pre-set alpha of 0.05). The resulting number of Orwin fail-safe was n = 4. In other words, only four potentially "missing" null effects, if added to the existing collection, would bring the observed combined Hedges' g++ to the "trivial" value of 0.1. These findings are largely in line with the overall observation of the research literature on online learning and distance education that finds this type of instruction is as good as but not necessarily better than more traditional face-to-face instructional interventions (see, for example, Bernard, Abrami, Lou et al., 2004). Naturally, the overall weighted average effect size in this collection of included meta-analyses is classified as small (in Cohen's terms, 1988), though statistically significant (p =.02), but only marginally stable.

3. ONE-STUDY-REMOVED SENSITIVITY ANALYSIS

One-study-removed sensitivity analysis identified as a potential outlier one relatively large effect of g++ = 0.777 that could constitute a "large leverage" effect size capable of disproportionately influencing the overall findings. Subsequently, the summary table that follows is organised to account for its presence/removal within both fixed and random analytical models (Table 15).

Table 15. Online learning overall weighted average effect size for achievement outcomes and associated heterogeneity statistics: Final data set

Population estimates	*k*	*g*++	SE	Lower 95th confidence interval	Upper 95th confidence interval
Online learning: With the ES from Roberts (2011)					
Random effects model	11	0.169*	0.07	0.03	0.31
Fixed effect model	11	0.130*	0.05	0.03	0.23
Heterogeneity analysis	Q_T = 16.60 (df = 10), p = .084, I^2 = 39.77				
Online learning: Without the ES from Roberts (2011)					
Random effects model	10	0.085	0.05	-0.01	0.18
Fixed effect model	10	0.085	0.05	-0.01	0.18
Heterogeneity analysis	Q_T = 4.16 (df = 9), p = .901, I^2 = 0.00				

* $p < 0.05$

Table 15 shows that, despite previously reported non-significant findings of the publication bias analysis, the stand-alone effect size of g++ = 0.777 is influential enough to impose at least two changes on the overall findings when this bias effect is removed from the distribution. While the resulting overall average effect size (random effects model) is statistically significant (p = .02) when this potential outlier is removed, it renders the resulting effect size non-significant (p = .09). Also, and no less importantly, the removal of this effect size substantially reduces the heterogeneity of the distribution. Non-significant to begin with (Q_{Total} = 16.60, p = .084), it drops dramatically to a negligible level of Q_{Total} = 4.16 (p = .901) and makes the results of both analytical models identical. It means that there is no unexplained variability left in the resulting distribution (I^2 = 0.00). Under these circumstances, any subsequent moderator variable analysis is meaningless. Indeed, the purpose of such analysis is to explore potential sources of systematic variation in effect sizes — with no unexplained variability in the distribution there is nothing to analyse. Even if the effect size in question is not removed, the distribution of the effect sizes in the Online Learning collection remains fairly homogeneous.

We tried to understand what made the meta-analysis by Roberts (2011) such a special case from a conceptual and methodological standpoint and if there is a justification (except for statistical estimates) for its exclusion from the Online

Learning collection of meta-analyses. First we noted that the meta-analysis focused exclusively on adult learning, which sets it apart from the other Online Learning meta-analyses as it is understood that adults differ from K–12 and early university-level students with regards to their self-regulation skills and their motivation to participate in, commit to and adequately manage online courses or training programmes. In addition, the meta-analysis made use of an atypical literature search and inclusion process where the author completed a new search and review process while also making use of two previously completed meta-analyses, namely Bernard et al. (2009) and Means et al. (2009). The author assumed that all studies published prior to 2005 were already included in Bernard et al. or Means et al., and hence only made use of the articles included in them, and then ran a new search from 2005 up to 2010. Such a literature search and review process presents a rather unorthodox approach that brings into question the systematicity of the meta-analysis. As such, and while not caught in the original review process, the Roberts meta-analysis constitutes a case that does not fully fit the criteria of systematicity. Hence, it is appropriate to remove it from the Online Learning collection of meta-analyses.

Blended Learning Collection

We followed the same standard sequence of procedures with the Blended Learning (including flipped classroom) collection of meta-analyses. The outcomes are given below.

Publication bias and potential outliers analyses

1. DUVAL AND TWEEDIE'S TRIM AND FILL ANALYSIS

Duval and Tweedie's trim and fill analysis and the related funnel plot are shown in Figure 6. According to the random effects model, the observed point estimate and the corresponding 95% confidence interval (in parentheses) is g++ = 0.470 ($CI_{L\text{-}U}$ = 0.32, 0.62). As in the Online Learning collection of meta-analyses, the trim and fill analytical routine and examination of the funnel plot found the distribution of the effect sizes in the Blended Learning collection to be quite balanced, with only one visibly outstanding large effect size. Also as with the Online Learning collection, the status and influence of this effect (g++ = 1.240, Shahnama et al., 2021) is further explored in the sensitivity analysis, and the aggregated data in the upcoming summary table report both overall effects — with and without this study. However, this analysis did not detect any publication bias and called for no adjustment. No imputation of potentially "missing" null effects is needed.

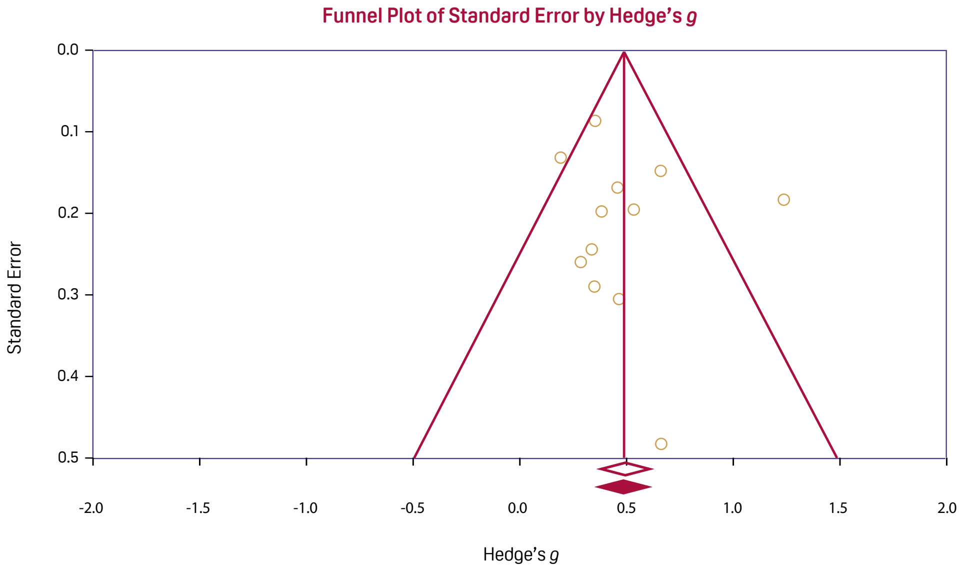

Figure 6. Funnel plot of achievement effects sizes for the Blended Learning collection of meta-analyses.

2. CLASSIC FAIL-SAFE ANALYSIS

Classic fail-safe analysis identified n = 275 as the number of potentially "missing" null effect studies (or over 20 per each included effect) required to render the observed overall weighted average effect size non-significant. The result of the Orwin fail-safe was n = 44 of potentially "missing" null effects needed for "trivialisation" of the observed Hedges' g++ — that is, to bring it to the value of 0.1. In other words, the overall weighted average effects size for the Blended Learning collection is not only significantly positive (moderate in terms of Cohen's classification, 1988) — it exceeds the effect of In-Class Technology Integration (though no direct statistical comparison between the two is possible) — but also robust (stable). The research literature on Blended Learning basically agrees (e.g., Bernard, Borokhovski, Schmid, Tamim & Abrami, 2014; Means et al., 2013).

3. ONE-STUDY-REMOVED SENSITIVITY ANALYSIS

One-study-removed sensitivity analysis did confirm that the outstanding effect size (g++ = 1.24) may indeed be an outlier capable of disproportionately affecting the overall outcomes of this second-order meta-analysis as reflected in the summary table below (Table 16), featuring both random and fixed analytical models.

Table 16. Blended learning overall weighted average effect size for achievement outcomes and associated heterogeneity statistics: Final data set

Population estimates	*k*	*g*++	SE	Lower 95th confidence interval	Upper 95th confidence interval
Blended learning: With the ES from Shahnama et al. (2021)					
Random effects model	13	0.470**	0.08	0.32	0.62
Fixed effect model	13	0.438**	0.05	0.35	0.53
Heterogeneity analysis	Q_T = 27.40 (df = 12), p = .007, I^2 = 56.20				
Blended learning: Without the ES from Shahnama et al. (2021)					
Random effects model	12	0.385**	0.05	0.29	0.48
Fixed effect model	12	0.385**	0.05	0.29	0.48
Heterogeneity analysis	Q_T = 7.34 (df = 11), p = .771, I^2 = 0.00				

** p < 0.01

Remarkably, the influence of one potential outlier in the Blended Learning collection mirrors that of the outlier in the Online Learning collection of effect sizes, at least with respect to the assessment of variability of the distribution. Its removal resulted in the significantly heterogeneous distribution of the effect sizes (Q_{Total} = 27.40, p = .007) turning into a perfectly homogeneous one (Q_{Total} = 7.34, p = .771) with a literally non-existing amount of unexplained variability (I^2 = 0.00), identical in both analytical models. The change to the magnitude of the overall weighted average effect size was not particularly dramatic: from g++ = 0.470 to g++ = 0.385, still highly statistically significant (p <.01) according to the random effects analytical model, and thus indicative of a marginally moderate ("small to moderate" according to some classifications) effect of blended learning on students' achievement.

The major difference between the Online Learning and Blended Learning collections of meta-analyses under review is that the latter could be further subjected to the moderator variable analyses if the detected outlier is kept within the distribution. Typically, the methodology of meta-analytical research suggests removing, winsorising (assigning less extreme values according to various statistical procedures) or explaining (a potentially reasonable option, especially in combination with either removing or winsorising) the outliers. We opted for a combination. First, we attempted to identify an explanation (at least a description) for the outstanding findings of the meta-analysis by Shahnama et al. (2021) in comparison with the rest of the collection. We also suggested interpreting the Blended Learning collection without this study but still ran some of the moderator variable analyses with the outlier retained. The rationale for this decision was that it might enable us to detect some tendencies without necessarily relying on the exact magnitude of the resulting effect sizes of the corresponding subgroups (levels of each of the assessed moderator variables).

A more in-depth review of the Shahnama et al. (2021) document revealed that from the 69 studies included in the meta-analyses, there were several effect sizes that might be considered potential outliers. More specifically, there were seven effect sizes in the 2–3 magnitude range and two in the 3–4 magnitude range, and one was 4.51 and another 6.32. Shahnama et al. conducted publication bias analysis, but no outlier analyses were completed. While the seven effect sizes in the 2–3 magnitude range might not have proved to be outliers, the 4.51 effect size is questionable, and the 6.32 effect size is a starkly obvious outlier in the provided funnel plot. As such, it is safe to consider that the effect size presented by Shahnama et al. is an overestimation based on the inclusion of a clear outlier, and hence it would be more appropriate to remove it from the Blended Learning collection of meta-analyses.

Though the moderator variable analyses were carried out with the Blended Learning collection of effect sizes (before removing the outlier), they all produced non-significant results. The comparison between the flipped classrooms and other forms of blended learning was of most interest for research and practice. Though the outcomes of this analysis (see Table 17, below) superficially support the flipped classroom instructions (g++ = 0.526 vs g++ = 0.397, respectively), this difference is not statistically significant: $Q_{Between}$ = 0.57, p = .450. Also, it is important to remember that the outlier was a part of this sub-collection, thus possibly overinflating its effect size, and so this finding can only be perceived as suggestive, not definitive.

Table 17. Moderator variable analysis (mixed effects model): Major function of technology revisited

Levels of moderator variables	Number of effects (k)	Category average ES (g++)	Lower 95th confidence interval	Upper 95th confidence interval	$Q_{Between}$
Flipped classrooms vs generic blended learning					
Flipped classrooms	7	0.526**	0.29	0.77	
Blended learning (generic)	6	0.368**	0.19	0.54	
Between groups: df = 1					1.10, p = .293

** p < .01

Discussion

The research questions that guided this second-order meta-analysis touched on various aspects of using technology in education. Data collected from 131 first-order meta-analyses that summarised relevant research should contribute to a better understanding of the various factors to consider (objectives, areas of application, strength, limitations, etc.) when integrating technology in teaching and learning and deliver some evidence-based answers about its role and effects. The discussion section below is built around questions pertaining to this topic.

1. What is the overall weighted average effect size (as the comprehensive point estimate of effectiveness) of technology use in classroom, online and blended learning as reflected in student achievement outcomes?

The question about the overall effects of educational technology on learning required summarising (aggregating) effect sizes extracted from included meta-analyses in three separate categories (as they are not fully compatible across educational settings, learners' mindsets and subsequently, the roles of educational technology).

In-Class Technology Integration

The overall effect of in-class technology integration was $g{+}{+} = 0.415$ ($k = 94$, after removal of "single region" meta-analyses). It is a low to moderate effect in terms of Cohen's (1988) categorisation. This means that, compared to the mean achievement score of the control condition (i.e., not using educational technology) of 50%, the mean of the experimental condition (i.e., teaching/learning with technology) would move to 61.5%, which constitutes an improvement of 11.5 percentage points. Its applied value would depend on the cost-effect analyses carried out by educational policymakers, administrators and practitioners for individual technologies in their corresponding educational settings — an estimated 11.5% of improved knowledge/skills may be worth a particular amount of investment (financial/infrastructural, human) under some circumstances (e.g., depending on the field of study, scope of application, specific educational objectives, etc.), but may be judged excessive under different circumstances. Moreover, taking into consideration the adjustment suggested by the trim and fill analyses, we may be dealing with a smaller effect of 0.347 (after statistical imputation of 17 potentially missing studies), which translates into only 9.7% improvement in the performance of an average "experimental learner" over that of a learner in the "control" group. However, it is reasonable to conclude that integrating technology into the educational process has the potential to positively impact learning achievement outcomes, possibly because of the increasing presence of technology in life outside of classrooms and its evolving functionality.

The distribution of effects in this category was also highly heterogeneous, ranging in magnitude from -0.03 — for using digital text to support reading comprehension in compulsory school students (Berkeley et al., 2015) — to 1.28 — for applying computer-assisted language learning (CALL) to improve quality of writing in adult second language learners (Xu et al., 2019). This heterogeneity suggests differential effects of learning with technology on student achievement. Potential systematic sources of this variability were explored in subsequent moderator variable analyses and will be discussed later.

Online Learning

Eleven meta-analyses initially included in this category (also with only studies with the international coverage retained) produced the overall weighted average effect size of $g_{++} = 0.169$ ($k = 11$), low in Cohen's terms and just marginally significant. Furthermore, our one-study-removed analyses identified a potential outlier, whose removal resulted in a noticeable drop in the overall effect to $g_{++} = 0.085$. This effect is obviously small, statistically not different from 0 and lower than what is typically referred to as "trivial" (0.10). In a sense, it repeats the overall findings of the most comprehensive meta-analysis in the field of distance education (Bernard, Abrami, Lou et al., 2004) that learning at a distance is "as good as" studying in-class (despite nearly 20 years having passed since it was published), at least in terms of students' achievement outcomes. Expressing it in terms of the percentile gain would mean 2.4% improvement, which may not justify the additional investment usually required for setting up an effective online learning environment. However, when alternatives are limited because of circumstances (e.g., access to quality teaching for students in remote areas, students' need for more flexible schedules, pandemic-related restrictions) this effect points to online learning as a viable option for education delivery.

However, educational practitioners must consider factors beyond simply achievement outcomes before deciding in favour of online learning. These factors include — but are not limited to — increased risk of social isolation and increased inequality among learners, as quality online learning requires access to infrastructure and technology itself, as well as to facilities (home- or community-based) that could provide interruption-protected learning environments; that is, conditions that vary substantially for learners from different socio-economic backgrounds. Perhaps even more important, as research (including our own) has repeatedly demonstrated, successful online learning relies on sound pedagogical principles and meaningful instructional design (Bernard, Borokhovski & Tamim, 2014). For more detailed discussion of potential pitfalls of hasty (emergency-driven) implementation of online learning, see Borokhovski et al. (2021).

Blended Learning

Thirteen meta-analyses initially fell into this category (again we only considered studies with international coverage). Together they produced the overall weighted average effect size of $g_{++} = 0.470$ ($k = 13$), rather moderate in Cohen's terms but statistically significant, further reduced to $g_{++} = 0.385$ (after the removal of one outlier identified by applying one-study-removed analytical routine). This effect returns us to the "low to moderate" category and translates into a 10.7% average improvement in academic performance of the experimental condition (in this case, blended learning) over that of the control condition (in this case, face-to-face instruction).

Though superior in learning achievements to online learning (i.e., relatively and not by direct comparison), blended learning may need more efforts on pedagogical design and planning. While the infrastructure (hardware, software and connectivity capabilities) required for delivering its distance component may be as sophisticated as that required for fully online learning, the in-class component of blended learning still requires access to auditoriums, lab facilities (to be maintained and serviced to the same degree as they would be for entirely face-to-face instruction) or both, and combining the two implies careful co-ordination and scheduling. Bele and Rugelj (2007) had high expectations not only of the potential of blended learning but also of the likelihood of its being widely implemented soon, and proclaimed this developing educational mode to be "the best of both worlds." As the multitude of research studies (including the first-order meta-analyses reviewed here) shows, the reality is not quite so simple. The "best of both worlds" results only emerge from "best of both worlds" efforts, which necessarily includes meaningful (i.e., adequately applied) pedagogy and well-thought-through instructional design. The latter assertion could be partly tested via future exploration of the Added Value studies, where both instructional conditions employ the same technology or delivery mode (either online or blended), but technology is supplemented (enhanced/supported) by some additional instructional quality, rooted in a pedagogical framework, instructional design solution or advanced technological functionality. Informed by such research evidence, our applied decisions about future directions and optimal strategies and conditions for employing blended learning should become more reliable. Meanwhile, to reiterate, the overall effect of blended learning on learning achievement outcomes is significant and exceeds (relatively, as a direct statistical comparison is not possible) the effects of in-class technology integration and, even more so, of online learning.

2. What moderator variables (methodological, substantive and contextual) influence learning with technology in these educational settings and to what extent?

Different kinds of moderator variables were addressed in the corresponding subsequent analyses primarily for the In-Class Technology Integration collection of included meta-analyses, as it was the only one heterogeneous enough to allow us to search for potential sources of systematic variation in effect sizes.

There was one exception made for the blended learning data set to estimate the differential effects of flipped classrooms and other forms of blended learning. The former turned out to be higher (though the difference was not statistically significant) than the latter: $g_{++} = 0.526$ and $g_{++} = 0.368$, respectively. It is possible that the difference is (at least partially) because the online component of flipped classrooms is less challenging to implement (see the brief discussion on page 8). For example, it could be more straightforward (possibly to the point of standardisation) to develop and deliver online mini-lectures prior to in-class active learning than more diverse activities in other forms of blended learning. This possibility may merit further attention, but as stated earlier, the difference based on the date of this second-order meta-analysis is not significant.

No other moderators were tested due to the relative homogeneity of the data set in the Blended Learning collection of meta-analyses.

With respect to in-class technology integration, the whole set of moderator variable analyses was conducted. We open the discussion of their results with the methodological moderator variables as the results of their analyses are directly linked to the response to the third research question. Only significant findings (or tendencies approaching levels of significance) are discussed here.

3. Does the methodology of primary meta-analyses under review matter?

Contrary to popular belief, meta-analysis, as the means for summarising findings of quantitative empirical research, is not protected from potential bias simply because it is more comprehensive. In reality, its scope (of accumulated information from multiple sources) makes meta-analysis more vulnerable to a threat of accumulation bias, as that bias may come not only from the inclusion of poorly conducted primary studies, but also from possible mistakes made by meta-analysts above and beyond inadequately rigorous research practices (see Bernard, 2014). This review and some of our previous works (e.g., Tamim et al., 2021) are not the first to imply a potential link between shortcomings in implementing a meta-analysis and potentially overinflating the magnitude of the estimated overall effect size (e.g., Cheung & Slavin, 2016; Pigott & Polanin, 2020). Findings from this current meta-analysis largely confirm this assumption.

Specifically, our review found that meta-analyses that systematically checked for methodological quality in the included primary studies (e.g., research design or dependency in data sets) and assessed the threat of introducing bias in their implementation (i.e., publication bias, outliers) produced significantly lower overall weighted average effect sizes than other meta-analyses. Similarly, comparisons between meta-analyses thoroughly exploring moderator variables and those in which such analyses were limited in scope or depth showed that the latter resulted in higher effects than the former. This difference, however, was not statistically significant. In summary, our answer to the third research question is unequivocally positive: the methodological quality of implementing meta-analytical research does matter, as it affects the findings. Our recent study identified which procedural components of a meta-analysis are most likely to lead to the effect size being overestimated if they are not implemented with adequate methodological quality (Tamim et al., 2021):

1. defining and contextualising the research problem
2. operationally defining the experimental and control conditions to highlight the factor(s) that underlies the comparison between the two
3. specifying outcome measures on which the effect size calculations are based
4. establishing the validity of the included primary research reports
5. checking for and maintaining independence of effect sizes

6. properly dealing with outliers
7. performing meaningful moderator variable analysis

4. Does the major function (objective) of technology use affect learning outcomes?

We coded for and tested the following levels of this moderator variable:

- multipurpose (all-encompassing, either specified or not specified — e.g., generic “computer-assisted instruction”) use of educational technology
- two types of cognitive support — deep learning (e.g., modelling complex social or biological processes in computer simulations) and distributed cognition (e.g., SPSS for performing statistical analyses or MS Word for delegating spell-checking to its processor to free up attentional resources to concentrate on writing)
- gaming (various forms of educational games)
- interaction/collaboration among learners (e.g., computer-mediated communication or computer-supported collaborative learning)
- content presentation (e.g., graphic illustrations to the text-based study materials)

None of these functions emerged as significantly more effective in the series of moderator variable analyses. At the same time, most of the distributions of the effect sizes within these sub-collections were significantly heterogeneous (i.e., with higher and lower effects for the same functionality). Perhaps focal functionality of educational technology by itself is not a driving force of its effectiveness, but how adequately it is applied for achieving specific educational goals is.

5. Is learning in the fields of study (subject matter disciplines) differentially successful when supported by educational technology?

Our analysis of this moderator variable found that among subject matters, language learning, and especially second language acquisition, benefits the most from support from educational technology ($g++ = 0.699$), also one of the most frequent forms of technology application focused on a specific discipline ($k = 17$). At this point, it is difficult to pinpoint why this is. It could be that earlier success in employing technology (e.g., hypertext glosses, as summarised in Yun, 2011) prompted follow-up research in the same field of study but with different types of technology (e.g., Mahdi, 2018, summarising data on mobile computer-assisted language learning). Or perhaps the effectiveness of technology to support language learning (often for developing vocabulary knowledge) varied across different tools, applications and categories of learners, but was consistent across time frames because it is such a good match for the task at hand, or the instructional design for this type of technology employment found especially successful solutions, or both. Other subjects (except first language learning, which is also associated with quite a high effect size) produced visibly lower (though statistically significant) effects. That applies to all-encompassing (across disciplines) meta-analyses as well.

6. Is educational technology differentially effective for learners at different academic grade levels/ages?

We found very little difference in effect sizes across coded grade levels (age groups) of learners. They are all clustered around an overall unadjusted mean of 0.415 with very minor deviations. Most likely, the selection of technological tools and applications was optimal — that is, adequate for learners of various ages and academic levels — eliciting on average similar responses in terms of learning achievements.

7. How effective was technology use for learning over time?

We once observed (Schmid et al., 2014) that the regression line reflecting the distribution of effect sizes over time is flat. We offered a possible, albeit speculative, explanation: That educational technology, as a reflection of rapid and escalating technological advances in general, simply does not have enough time to prove itself in rigorous empirical research. By the time it is widely enough recognised and accepted by educational practitioners (which necessarily includes in-place infrastructure and training and technical support for teachers, plus instructional design tailored to a particular educational tool/application) and ready to be extensively but rigorously studied by educational researchers, a new technology touted as "cutting edge," "revolutionary" and so on attracts stakeholders' attention. This hypothetical cycle of "emergence of — fascination with — adoption for learning — research into effectiveness of use — switching to a new, more fashionable 'shiny object'" takes on average three to five years so that the corresponding research snapshots always capture the same level of effectiveness humanly achievable in a given time, despite all the (perceived or/and real) advances and wonders of the technology in question.

However, this second-order meta-analysis detected a slight change in the pattern of results. The regression line "effect size by date" is still relatively flat. Its slope is not pronounced, but it gained in significance in comparison with Schmid et al. (2014): $p = .012$, both for publication date and publication coverage regression analyses. It appears that more recent meta-analyses tend to report slightly higher average effect sizes. Testing publication date as a categorical variable confirms that empirical studies summarised in the meta-analyses published after 2010 on average produced significantly ($p = .004$) higher effects ($g{+}{+} = 0.452$) than the studies included in earlier (i.e., prior to 2011) meta-analyses ($g{+}{+}= 0.304$). Though there is a substantial overlap in coverage between the two collections, it appears that with time, in-class technology integration gradually gains in effectiveness.

8. What additional aspects and emerging directions of research on technology integration in education deserve closer examination in subsequent systematic reviews?

In response to this question, we would like to briefly outline what directions in studying technology effects on learning appear to be of special interest for future research.

This report contains a subsection on meta-analyses that summarise research on special needs students and added value studies of educational technology.

First, while technology use in a regular classroom has to date been a choice on the part of educational practitioners and participation in online (technology-dependent) learning largely students' choice (at least until COVID-19 disease control measures were introduced), technology in the service of learners with special needs may indeed be a game changer or, at least, an essential instrument for adequately addressing these special needs. We need to continue rigorously researching what works best, or better, for different categories of learners to provide much-needed targeted support.

Second, as we argued in the case of distance education, very little insight could be gained through studies comparing technology-saturated (enabled/supported) learning environments with technology-free ones (which are rapidly becoming obsolete even in poorer regions of the world). Such comparisons would not be able to explore and explain what technologies work better or worse under what circumstances and why. Only comparing different uses (first and foremost, in terms of adopted pedagogical frameworks and instructional design feature) of the same technological tools or applications would contribute to our understanding of the true potential of educational technology and guide its meaningful application.

In addition, we believe that more rigorous empirical research into emerging educational technologies is needed, if only to shorten the duration of the fashion-driven cycle described earlier. It is important to verify the advantages and identify the shortcomings of new trends before interest in the former dwindles and too many mistakes are made because of the latter. Of course, research should continue to address moderator variables in the studies of well-established educational technologies by going beyond basic demographics and contextual features towards in-depth examination of more nuanced characteristics of what instructional purpose the technology under consideration serves and what pedagogical principles underlie its use.

As it stands now, the research has clearly established the positive, though low to moderate, effects of integrating information and communication technologies into the educational process on students' learning outcomes. More importantly, it has revealed that these effects may vary substantially, depending on a number of methodological, instructional and demographic variables. Specifically, what matters the most is pedagogical underpinnings of the interventions in question. Instructional design that carefully crafts meaningful use of educational technology seems to outweigh the advancements of technology itself. In other words, the focus on how educational technology is used is more important than if and what technology to use.

Conclusion

As technology becomes an increasingly commonplace tool in educational settings of all descriptions around the world there will be an increasing need to monitor its effectiveness and assess its benefits and drawbacks. New research will emerge in tandem with and as a result of the increasing use of technology — and, as we have seen here, that research will come complete with its own biases. If we are to stay abreast of developments in technology, the use of technology in an educational setting and the effectiveness of using technology in an educational setting, it is crucial that robust independent literature reviews — including both first-order and second-order meta-analyses — are conducted regularly. Follow-up meta-analyses should also bring us closer to identifying the elusive "something" that students gain from learning in a hybrid classroom-based/online learning environment. That in turn will help us to understand the optimal balance between in-class and online learning and to work towards establishing best practices, strategies and techniques for technology-supported hybrid learning.

References

Studies in the meta-analysis indicated by an asterisk (*)

*Abrami, P. C., Lysenko, L., & Borokhovski, E. (2020). The effects of ABRACADABRA on reading outcomes: An updated meta-analysis and landscape review of applied field research. *Journal of Computer Assisted Learning, 36*(3), 260-279. https://doi.org/10.1111/jcal.12417

*Allen, M., Mabry, E., Mattrey, M., Bourhis, J., Titsworth, S., & Burrell, N. (2004). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. *Journal of Communication, 54*(3), 402-420. https://doi.org/10.1111/j.1460-2466.2004.tb02636.x

Al-Jewair, T. S., Amir, A., Sunjay, S., & Prakesh, S. S. (2009). Computer-assisted learning in orthodontic education: A systematic review and meta-analysis. *Journal of Dental Education, 73*(6), 730-739. https://doi.org/10.1002/j.0022-0337.2009.73.6.tb04752.x

Al-Wasy, B. Q. (2020). The effectiveness of integrating technology in EFL/ESL writing: A meta-analysis. *Interactive Technology and Smart Education, 17*(4), 435-454. https://doi.org/10.1108/ITSE-03-2020-0033

Azorín, C. (2020). Beyond COVID-19 supernova. Is another education coming? *Journal of Professional Capital and Community*, *5*(3/4), 381-390. https://doi.org/10.1108/JPCC-05-2020-0019

*Bayraktar, S. (2002). A meta-analysis of the effectiveness of computer-assisted instruction in science education. *Journal of Research on Technology in Education, 34*(2), 173-188. https://doi.org/10.1080/15391523.2001.10782344

Bele, J. L., & Rugelj, J. (2007). Blended learning — an opportunity to take the best of both worlds. *International Journal of Emerging Technologies in Learning, 2*(4). https://www.researchgate.net/publication/26474056_Blended_Learning_-_An_Opportunity_to_Take_the_Best_of_Both_Worlds

*Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309-344. https://doi.org/10.3102/0034654316670999

*Belland, B. R., Walker, A. E., Olsen, M. W., & Leary, H. (2015). A pilot meta-analysis of computer-based scaffolding in STEM education. *Educational Technology & Society, 18*(1), 183-197. https://eric.ed.gov/?id=EJ1062484

*Benavides-Varela, S., Zandonella Callegher, C., Fagiolini, B., Leo, I., Altoè, G., & Lucangeli, D. (2020). Effectiveness of digital-based interventions for children with mathematical learning difficulties: A meta-analysis. *Computers & Education, 157*, 103953. https://doi.org/10.1016/j.compedu.2020.103953

*Berkeley, S., Kurz, L. A., Boykin, A., & Evmenova, A. S. (2015). Improving reading comprehension using digital text: A meta-analysis of interventions. *International Journal for Research in Learning Disabilities, 2*(2), 18-43.

Bernard, R. M. (2014). Things I have learned about meta-analysis since 1990: Reducing bias in search of "the big picture." *Canadian Journal of Learning and Technologies, 40*(3). https://files.eric.ed.gov/fulltext/EJ1038400.pdf

*Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., & Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439. https://doi.org/10.3102/00346543074003379

Bernard, R. M., Abrami, P. C., Wade, A., Borokhovski, E., Lou, Y. (2004). *The effects of synchronous and asynchronous distance education: A meta-analytical assessment of Simonson's "equivalency theory."* Association for Educational Communications and Technology. https://eric.ed.gov/?id=ED485078

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

Bernard, R. M., Borokhovski, E., Schmid, R. F., & Tamim, R. M. (2014). An exploration of bias in meta-analysis: The case of technology integration research in higher education. *Journal of Computing in Higher Education, 26*, 183-209. https://doi.org/10.1007/s12528-014-9084-z

Bernard, R. M., Borokhovski, E., Schmid, R. F., & Tamim, R. M. (2018). Gauging the effectiveness of educational technology integration in education: What the best-quality meta-analyses tell us. In M. Spector, B. Lockee, & M. Childress (Eds.), *Learning, design, and technology* (pp. 1-25). Springer. https://doi.org/10.1007/978-3-319-17727-4_109-1

*Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education, 26*(1), 87-122. https://doi.org/10.1007/s12528-013-9077-3

Bernard, R. M., Borokhovski, E., Tamim, & R. M. (2014). Detecting bias in meta-analyses of distance education research: big picture we can rely on. *Distance Education 35*(3), 271-293. https://doi.org/10.1080/01587919.2015.957433

Bethel, E. C., & Bernard, R. M. (2010). Developments and trends in synthesizing diverse forms of evidence: Beyond comparisons between distance education and classroom instruction. *Distance Education, 31*(3), 231-256. https://doi.org/10.1080/01587919.2010.513950

Bishop, J. L., & Verleger, M. A. (2013, June). *The flipped classroom: A survey of the research* [Paper presentation]. 120th ASEE annual conference & exposition, Atlanta, GA. https://www.researchgate.net/publication/285935974_The_flipped_classroom_A_survey_of_the_research

Blok, H., Oostdam, R., & Otter, M. E. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research, 72*(1), 101-130. https://doi.org/10.3102/00346543072001101

*Bogomolova, K., Hierck, B. P., Looijen, A. E. M., Pilon, J. N. M., Putter, H., Wainman, B., Hovius, S. E. R., & Hage, J. A. (2021). Stereoscopic three-dimensional visualisation technology in anatomy learning: A meta-analysis. *Medical Education, 55*(3), 317-327. https://doi.org/10.1111/medu.14352

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97-111. https://doi.org/10.1002/jrsm.12

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). *Comprehensive Meta-Analysis.* In (Version 3.3.070) Biostat. https://www.meta-analysis.com/pages/new_v3.php?cart=B2NM6705381

Borokhovski, E., Bernard, R. M., Tamim, R. M., Schmid, R. F., & Sokolovskaya, A. (2016). Technology-supported student interaction in post-secondary education: A meta-analysis of designed versus contextual treatments. *Computers & Education, 96*, 15-28. https://doi.org/10.1016/j.compedu.2015.11.004

Borokhovski, E., Pickup, D., & Tamim, R. M. (2021). "Emergency distance education" model: How normal could the projected new normal be? In V. Kankhva (Ed.), *E3S Web of Conferences, Ural Environmental Science Forum "Sustainable Development of Industrial Region" (UESF-2021)*. https://doi.org/10.1051/e3sconf/202125807072

Borokhovski, E., Tamim, R. M., Bernard, R. M., Abrami, P. C., & Sokolovskaya, A. (2012). Are contextual and design student-student interaction treatments equally effective in distance education? A follow-up meta-analysis of comparative empirical studies. *Distance Education, 33*(3), 311-329. https://doi.org/10.1080/01587919.2012.723162

Burtson, J., & Arispe, K. (2018). Looking for a needle in a haystack: CALL and advanced language proficiency. *CALICO Journal, 35*(1), 77-102. https://doi.org/10.1558/cj.31594

*Byun, J., & Joung, E. (2018). Digital game-based learning for K–12 mathematics education: A meta-analysis. *School Science & Mathematics, 118*(3/4), 113-126. https://doi.org/10.1111/ssm.12271

Carpenter, C. R., & Greenhill, L. P. (1995). *An investigation of closed circuit television for teaching university courses: Report number one.* The Pennsylvania State University.

*Castillo-Manzano, J. I., Castro-Nuño, M., López-Valpuesta, L., Sanz-Díaz, M. T., & Yñiguez, R. (2016). Measuring the effect of ARS on academic performance: A global meta-analysis. *Computers & Education, 96*, 109-121. https://doi.org/10.1016/j.compedu.2016.02.007

*Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review, 33*(3), 989-1015. https://doi.org/http://dx.doi.org/10.1007/s10648-020-09587-1

*Cavanaugh, C., Gillan, K. J., Kromrey, J., Hess, M., Blomeyer, R., & North Central Regional Educational Lab. (2004). *The effects of distance education on k-12 student outcomes: A meta-analysis.* Learning Point Associates / North Central Regional Educational Laboratory (NCREL). http://files.eric.ed.gov/fulltext/ED489533.pdf

*Chan, K. K., & Leung, S. W. (2014). Dynamic geometry software improves mathematical achievement: Systematic review and meta-analysis. *Journal of Educational Computing Research, 51*(3), 311-325. https://journals.sagepub.com/doi/abs/10.2190/EC.51.3.c?journalCode=jeca

*Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education, 105*, 14-30. https://doi.org/10.1016/j.compedu.2016.11.005

Chen, C.-H., Shih, C.-C., & Law, V. (2020). The effects of competition in digital game-based learning (DGBL): A meta-analysis. *Educational Technology, Research and Development, 68*(4), 1855-1873. https://doi.org/10.1007/s11423-020-09794-1

*Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research, 88*(6), 799-843. https://doi.org/10.3102/0034654318791584

*Chen, M.-H., Tseng, W.-T., & Hsiao, T.-Y. (2018). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology, 49*(1), 69-77. https://doi.org/http://dx.doi.org/10.1111/bjet.12526

*Chen, Z., Chen, W., Jia, J., & An, H. (2020). The effects of using mobile devices on language learning: A meta-analysis. *Educational Technology Research & Development, 68*(4), 1769-1789. https://doi.org/10.1007/s11423-020-09801-5

*Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research and Development, 67*(4), 793-824. https://doi.org/10.1007/s11423-018-9633-7

*Cheung, A. C. K., & Slavin, R. E. (2011). *The effectiveness of education technology for enhancing reading achievement: A meta-analysis.* Center for Research and Reform in Education. https://eric.ed.gov/?id=ED527572

*Cheung, A. C. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review, 7*(3), 198-215. https://doi.org/10.1016/j.edurev.2012.05.002

*Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in k-12 classrooms: A meta-analysis. *Educational Research Review, 9*, 88-113. https://doi.org/10.1016/j.edurev.2013.01.001

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283-292. https://doi.org/10.3102/0013189X16656615

*Chien, Y.-T., Chang, Y.-H., & Chang, C.-Y. (2016). Do we click in the right way? A meta-analytic review of clicker-integrated instruction. *Educational Research Review, 17*, 1-18. https://doi.org/10.1016/j.edurev.2015.10.003

*Chiu, Y.-H. (2013). Computer-assisted second language vocabulary instruction: A meta-analysis. *British Journal of Educational Technology, 44*(2). https://doi.org/10.1111/j.1467-8535.2012.01342.x

*Cho, K., Lee, S., Joo, M.-H., & Becker, B. J. (2018). The effects of using mobile devices on student achievement in language learning: A meta-analysis. *Education Sciences, 8*(3), 105. https://doi.org/10.3390/educsci8030105

*Christmann, E. P., & Badgett, J. L. (2000). The comparative effectiveness of CAI on collegiate academic performance. *Journal of Computing in Higher Education, 11*(2), 91-103. https://eric.ed.gov/?id=EJ613263

*Christmann, E. P., & Badgett, J. L. (2003). A meta-analytic comparison of the effects of computer-assisted instruction on elementary students' academic achievement. *Information Technology in Childhood Education Annual, 2003*(1), 91-104. https://eric.ed.gov/?id=EJ823484

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-449. https://doi.org/10.3102/00346543053004445

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J., & Montori, V. M. (2008). Internet-based learning in the health professions: A meta-analysis. *JAMA, 300*(10), 1181-1196. https://doi.org/10.1001/jama.300.10.1181

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach.* SAGE.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* Russell Sage Foundation.

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*(6), 446-462. https://doi.org/10.1037/a0027119

*Costa, J. M., & Miranda, G. L. (2017). Relation between Alice software and programming learning: A systematic review of the literature and meta-analysis. *British Journal of Educational Technology, 48*(6), 1464-1474. https://doi.org/10.1111/bjet.12496

*Cromley, J. G., Du, Y., & Dane, A. P. (2020). Drawing-to-learn: Does meta-analysis show differences between technology-based drawing and paper-and-pencil drawing? *Journal of Science Education and Technology, 29*(2), 216-229. https://doi.org/10.1007/s10956-019-09807-6

Daniel, J. (2020). Education and the COVID-19 pandemic. *Prospects*, *49*, 91-96. https://doi.org/10.1007/s11125-020-09464-3

Darabi, A., Liang, X., Suryavanshi, R., & Yurekli, H. (2013). Effectiveness of online discussion strategies: A meta-analysis. *American Journal of Distance Education, 27*(4), 228-241. https://doi.org/10.1080/08923647.2013.837651

Doo, M. Y., Bonk, C. J., & Heo, H. (2020). A meta-analysis of scaffolding effects in online learning in higher education. *International Review of Research in Open and Distributed Learning, 21*(3), 60-80. https://eric.ed.gov/?id=EJ1267049

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

*Fang, Y., Ren, Z., Hu, X., & Graesser, A. C. (2019). A meta-analysis of the effectiveness of ALEKS on learning. *Educational Psychology, 39*(10), 1278-1292. https://doi.org/10.1080/01443410.2018.1495829

*Fengfeng, K., Huifen, L., Yu-Hui, C., & Dwyer, F. (2006). Effects of animation on multi-level learning outcomes for learners with different characteristics: A meta-analytic assessment and interpretation. *Journal of Visual Literacy, 26*(1), 15-40. https://doi.org/10.1080/23796529.2006.11674630

Fichten, C., Pickup, D., Asunsion, J., Jorgensen, M., Vo, C., Legault, A., & Libman, E. (2021). State of the research on artificial intelligence based apps for post-secondary students with disabilities. *Exceptionally Education International, 31*(1). https://doi.org/10.5206/eei.v31i1.14089

*Gegenfurtner, A., & Ebner, C. (2019). Webinars in higher education and professional training: A meta-analysis and systematic review of randomized controlled trials. *Educational Research Review, 28*, 100293. https://www.researchgate.net/publication/336343101_Webinars_in_Higher_Education_and_Professional_Training_A_Meta-Analysis_and_Systematic_Review_of_Randomized_Controlled_Trials

Gerard, L., Matuk, C., McElhaney, K., & Linn, M. C. (2015). Automated, adaptive guidance for K-12 education. *Educational Research Review, 15*, 41-58. https://doi.org/10.1016/j.edurev.2015.04.001

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Sage.

*Grgurovic, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL, 25*(2), 165-198. https://doi.org/10.1017/S0958344013000013

Gunter, G. A. (2012). Digital booktalk: Creating a community of avid readers, one video at a time. *Computers in the Schools, 29*(1-2), 135-156. https://doi.org/10.1080/07380569.2012.651426

Guo, Y. R., & Goh, D. H.-L. (2015). Affect in embodied pedagogical agents: Meta-analytic review. *Journal of Educational Computing Research, 53*(1), 124-149. https://doi.org/10.1177/0735633115588774

*Guo, D., Zhang, S., Wright, K. L., & McTigue, E. M. (2020). Do you get the picture? A meta-analysis of the effect of graphics on reading comprehension. *AERA Open, 6*(1). https://doi.org/10.1177/2332858420901696

*Hassan Saleh, M., & Ahmed Abdulateef Al, K. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education, 7*(3), 733-753. https://doi.org/http://dx.doi.org/10.1002/rev3.3165

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge. https://doi.org/10.4324/9780203887332

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis.* National Science Teachers Association.

Hew, K. F., Bai, S., Huang, W., Dawson, P., Du, J., Huang, G., Jia, C., & Thankrit, K. (2021). On the use of flipped classroom across various disciplines: Insights from a second-order meta-analysis. *Australasian Journal of Educational Technology, 37*(2), 132-151. https://doi.org/10.14742/ajet.6475

Hew, K. F., Bai, S., Huang, W., Du, J., Huang, G., Jia, C., & Khongjan, T. (2020). Does flipped classroom improve student cognitive and behavioral outcomes in STEM subjects? Evidence from a second-order meta-analysis and validation study. In S. Cheung, R. Li, K. Phusavat, N. Paoprasert, & L. Kwok (Eds.), *ICBL 2020: Blended learning. Education in a smart learning environment* (pp. 264-275). Springer. https://doi.org/10.1007/978-3-030-51968-1_22

*Higgins, K., Huscroft-D'Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research, 57*(2), 283-319. https://doi.org/10.1177/0735633117748416

*Holmes, A. B. (2013). *Effects of manipulative use on pk-12 mathematics achievement: A meta-analysis.* Society for Research on Educational Effectiveness. http://files.eric.ed.gov/fulltext/ED563072.pdf

*Howard, M. C., Gutworth, M. B., & Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior, 121*, 106808. https://doi.org/10.1016/j.chb.2021.106808

*Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers & Education, 94*, 102-119. https://doi.org/10.1016/j.compedu.2015.11.013

Jahng, N., Krug, D., & Zhang, Z. (2007). Student achievement in online distance education compared to face-to-face education. *European Journal of Open, Distance and E-Learning, 2007*(1). https://old.eurodl.org/?p=archives&sp=full&year=2007&halfyear=i&article=253

Jonassen, D. H. (1995). Computers as cognitive tools: Learning *with* technology, not *from* technology. *Journal of Computing in Higher Education, 6*, 40. https://doi.org/10.1007/BF02941038

Kim, J., & Kimm, C. H. (2017). Functional technology for individuals with intellectual disabilities: Meta-analysis of mobile device-based interventions. *Journal of Special Education Apprenticeship, 6*(1). https://eric.ed.gov/?id=EJ1149575

*Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for stem education: Bayesian meta-analysis. *Educational Psychology Review, 30*(2), 397-429. https://doi.org/10.1007/s10648-017-9419-1

Korunka, C., & Hoonakker, P. (2014). The future of ICT and quality of working life: Challenges, benefits, and risks. In C. Korunka & P. Hoonakker (Eds.), *The impact of ICT on quality of working life* (pp. 205-219). Springer. https://doi.org/10.1007/978-94-017-8854-0_13

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*(2), 179-211. https://doi.org/10.3102/00346543061002179

*Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42-78. https://doi.org/10.3102/0034654315581420

*Kunkel, A. K. (2015). *The effects of computer-assisted instruction in reading: A meta-analysis* (Publication No. 3727817) University of Minnesota. ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Låg, T., & Sæle, R. G. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-analysis. *AERA Open, 5*(3). https://doi.org/10.1177/2332858419870489

Lan, Y.-C., Lo, Y.-L., & Hsu, Y.-S. (2014). The effects of meta-cognitive instruction on students' reading comprehension in computerized reading contexts: A quantitative meta-analysis. *Educational Technology & Society, 17*(4), 186-202. https://eric.ed.gov/?id=EJ1045530

*Larwin, K., & Larwin, D. (2011). A meta-analysis examining the impact of computer-assisted instruction on postsecondary statistics education: 40 years of research. *Journal of Research on Technology in Education, 43*(3), 253-278. http://files.eric.ed.gov/fulltext/EJ918907.pdf

*Li, C., He, J., Yuan, C., Chen, B., & Sun, Z. (2019). The effects of blended learning on knowledge, skills, and satisfaction in nursing students: A meta-analysis. *Nurse Education Today, 82*, 51-57. https://doi.org/10.1016/j.nedt.2019.08.004

*Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22*(3), 215-243. https://doi.org/10.1007/s10648-010-9125-8

*Lin, H. (2015). A meta-synthesis of empirical research on the effectiveness of computer-mediated communication (CMC) in SLA. *Language Learning & Technology, 19*(2), 85-117. https://doi.org/10125/44419

*Lin, J.-J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning, 32*(8), 878-919. https://doi.org/10.1080/09588221.2018.1541359

*Lin, W.-C., Huang, H.-T., & Liou, H.-C. (2013). The effects of text-based SCMC on SLA: A meta analysis. *Language Learning and Technology, 17*(2), 123-142. https://www.researchgate.net/publication/285893659_The_effects_of_text-based_scmc_on_SLA_A_meta_analysis

*Little, C. W., Clark, J. C., Tani, N. E., & Connor, C. M. (2018). Improving writing skills through technology-based instruction: A meta-analysis. *Review of Education, 6*(2), 183-201. https://doi.org/10.1002/rev3.3114

Liu, Q., Peng, W., Zhang, F., Hu, R., Li, Y., & Yan, W. (2016). The effectiveness of blended learning in health professions: Systematic review and meta-analysis. *Journal of Medical Internet Research, 18*(1), e2. https://doi.org/10.2196/jmir.4807

*Lo, C. K., & Hew, K. F. (2019). The impact of flipped classrooms on student achievement in engineering education: A meta-analysis of 10 years of research. *Journal of Engineering Education, 108*(4), 523-546. https://doi.org/10.1002/jee.20293

Lou, Y., Bernard, R. M., & Abrami, P. C. (2006). Media and pedagogy in undergraduate distance education: A theory-based meta-analysis of empirical literature. *Educational Technology Research and Development, 54*(2), 141-176. https://doi.org/10.1007/s11423-006-8252-x

Lyngdoh, S. N., & Bene, K. (2018). The effect of video modeling on preschoolers' learning who are diagnosed with autism spectrum disorder: A meta-analysis. *Journal of the American Academy of Special Education Professionals*, Winter, 82-102. https://eric.ed.gov/?id=EJ1254605

*Mahdi, H. S. (2018). Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of Educational Computing Research, 56*(1), 134-154. https://doi.org/10.1177/0735633117698826

McElhaney, K. W., Chang, H.-Y., Chiu, J. L., & Linn, M. C. (2015). Evidence for effective uses of dynamic visualisations in science curriculum materials. *Studies in Science Education, 51*(1), 49-85. https://doi.org/10.1080/03057267.2014.984506

*McTigue, E. M., Solheim, O. J., Zimmer, W. K., & Uppstad, P. H. (2020). Critically reviewing GraphoGame across the world: Recommendations and cautions for research and implementation of computer-assisted instruction for word-reading acquisition. *Reading Research Quarterly, 55*(1), 45-73. https://doi.org/10.1002/rrq.256

*Means, B., Toyama, Y., Murphy, R. F., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 115*(3). https://journals.sagepub.com/doi/10.1177/016146811311500307

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* US Department of Education.

*Mei-Mei, C. (2019). Effects of technology-enhanced language learning on second language acquisition: A meta-analysis. *Journal of Educational Technology & Society, 22*(4). https://eric.ed.gov/?id=EJ1250441

*Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education, 70*, 29-40. https://doi.org/10.1016/j.compedu.2013.07.033

*Moran, J., Ferdig, R. E., Pearson, P. D., Wardrop, J., & Blomeyer, R. L., Jr. (2008). Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. *Journal of Literacy Research, 40*(1), 6-58. https://doi.org/10.1080/10862960802070483

*Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2021). Video improves learning in higher education: A systematic review. *Review of Educational Research, 91*(2), 204-236. https://doi.org/10.3102/0034654321990713

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics, 8*(2), 157-159. https://doi.org/10.3102/10769986008002157

*Özdemir, M., Şahin, C., Arcagok, S., & Demir, M. K. (2018). The effect of augmented reality applications in the learning process: A meta-analysis study. *Eurasian Journal of Educational Research* (74), 165-186. https://www.researchgate.net/publication/324210750_The_Effect_of_Augmented_Reality_Applications_in_the_Learning_Process_A_Meta-Analysis_Study

Page, M. J., Boutron, I., Shamseer, L., Brennan, S. E., Grimshaw, J. M., Li, T., McDonald, S., Thomas, J., & Whiting, P. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

*Peng, H., Jager, S., & Wander, L. (2021). Narrative review and meta-analysis of MALL research on L2 skills. *ReCALL: The Journal of EUROCALL, 33*(3), 278-295. https://doi.org/http://dx.doi.org/10.1017/S0958344020000221

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research, 28*(3), 450-461. https://doi.org/10.1086/323732

Pigott, T. D. (2012). *Advances in meta-analysis.* Springer Science+Business Media, LLC, New York, https://doi.org/10.1007/978-1-4614-2278-5

Pigott, T. D., & Polanin, J. R. (2020). Methodological guidance paper: High-quality meta-analysis in a systematic review. *Review of Educational Research, 90*(1), 24-46. https://doi.org/10.3102/0034654319877153

*Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 15*(1), 5-47. https://doi.org/10.1007/s11412-020-09316-4

*Rahmati, J., Izadpanah, S., & Shahnavaz, A. (2021). A meta-analysis on educational technology in English language teaching. *Language Testing in Asia, 11*(1), 1-20. https://doi.org/10.1186/s40468-021-00121-w

*Rakes, C. R., Ronau, R. N., Bush, S. B., Driskell, S. O., Niess, M. L., & Pugalee, D. K. (2020). Mathematics achievement and orientation: A systematic review and meta-analysis of education technology. *Educational Research Review, 31*, 100337. https://doi.org/10.1016/j.edurev.2020.100337

*Ran, H., Kasli, M., & Secada, W. G. (2021). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in k-12 classrooms. *Journal of Educational Computing Research, 59*(1), 119-153. https://doi.org/10.1177/0735633120952063

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ, 342*. https://doi.org/10.1136/bmj.d549

*Roberts, R. M. (2011). *Best instructional practices for distance education: A meta-analysis* (Publication No. 3475479) University of Nevada, Las Vegas. ProQuest Central; ProQuest Dissertations & Theses Global. Ann Arbor.

*Saad Mohamed, A. F. (2020). Feedback in computer-assisted language learning: A meta-analysis. *TESL-EJ, 24*(2). https://www.tesl-ej.org/wordpress/issues/volume24/ej94/ej94a4/

*Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin, 144*(2), 111-139. https://doi.org/10.1037/bul0000139

Sanders, M., & George, A. (2017). Viewing the changing world of educational technology from a different perspective: Present realities, past lessons, and future possibilities. *Education and Information Technologies, 22*, 2915-2933. https://doi.org/10.1007/s10639-017-9604-3

*Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2014). Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on Learning Technologies, 7*(1), 38-56. https://doi.org/10.1109/TLT.2013.37

*Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., Wade, C. A., & Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers & Education, 72*, 271-291. https://doi.org/10.1016/j.compedu.2013.11.002

*Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1-39. https://doi.org/10.2190/EC.49.1.a

Seo, Y.-J., & Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Computers & Education, 53*(3), 913-928. https://doi.org/10.1016/j.compedu.2009.05.002

*Shachar, M., & Neumann, Y. (2010). Twenty years of research on the academic performance differences between traditional and distance learning: Summative meta-analysis and trend examination. *Journal of Online Learning and Teaching, 6*(2), 318. https://www.researchgate.net/publication/242732581_Twenty_Years_of_Research_on_the_Academic_Performance_Differences_Between_Traditional_and_Distance_Learning_Summative_Meta-Analysis_and_Trend_Examination

*Shahnama, M., Ghonsooly, B., & Shirvan, M. E. (2021). A meta-analysis of relative effectiveness of flipped learning in English as second/foreign language research. *Educational Technology, Research and Development, 69*(3), 1355-1386. https://doi.org/http://dx.doi.org/10.1007/s11423-021-09996-1

*Sharifi, M., Rostami AbuSaeedi, A., Jafarigohar, M., & Zandi, B. (2018). Retrospect and prospect of computer assisted English language learning: A meta-analysis of the empirical literature. *Computer Assisted Language Learning, 31*(4), 413-436. https://doi.org/10.1080/09588221.2017.1412325

*Shi, Y., Zhang, J., Yang, H., & Harrison Hao, Y. (2021). Effects of interactive whiteboard-based instruction on students' cognitive learning outcomes: A meta-analysis. *Interactive Learning Environments, 29*(2), 283-300. https://doi.org/http://dx.doi.org/10.1080/10494820.2020.1769683

Simonson, M. (2015). Hooray! Or, here we go again! *Distance Learning, 12*(1), 45-46. https://www.proquest.com/docview/230741057

*Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*(2), 489-528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

*Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology, 59*(3), 623-664. https://doi.org/10.1111/j.1744-6570.2006.00049.x

*Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427-515. https://doi.org/http://dx.doi.org/10.3102/0034654308317473

*Sokolowski, A., Li, Y., & Willson, V. (2015). The effects of using exploratory computerized environments in grades 1 to 8 mathematics: A meta-analysis of research. *International Journal of STEM Education, 2*. https://doi.org/10.1186/s40594-015-0022-z

*Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research, 81*(1), 97-128. https://doi.org/10.3102/0034654310378174

Spanjers, I. A. E., Könings, K. D., Leppink, J., Verstegen, D. M. L., de Jong, N., Czabanowska, K., & van Merriënboer, J. J. G. (2015). The promised land of blended learning: Quizzes as a moderator. *Educational Research Review, 15*, 59-74. https://doi.org/10.1016/j.edurev.2015.05.001

Sparkes, C. (2020). *Flipped classrooms versus traditional classrooms: A systematic review and meta-analysis of student achievement in higher education* [Doctoral dissertation]. Concordia University. Montreal, QC.

*Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on k-12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970-987. https://doi.org/10.1037/a0032447

*Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106*(2), 331-347. https://doi.org/10.1037/a0034752

*Sun, S., Else-Quest, N. M., Hodges, L. C., French, A. M., & Dowling, R. (2021). The effects of ALEKS on mathematics learning in k-12 and higher education: A meta-analysis. *Investigations in Mathematics Learning, 13*(3), 182-196. https://doi.org/10.1080/19477503.2021.1926194

*Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education, 94*, 252-275. https://doi.org/10.1016/j.compedu.2015.11.008

*Sung, Y.-T., Chang, K.-E., & Yang, J.-M. (2015). How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review, 16*, 68-84. https://doi.org/10.1016/j.edurev.2015.09.001

*Sung, Y.-T., Yang, J.-M., & Lee, H.-Y. (2017). The effects of mobile-computer-supported collaborative learning: Meta-analysis and critical synthesis. *Review of Educational Research, 87*(4), 768-805. https://doi.org/10.3102/0034654317704307

*Taj, I. H., Sulan, N. B., Sipra, M. A., & Ahmad, W. (2016). Impact of mobile assisted language learning (MALL) on EFL: A meta-analysis. *Advances in Language and Literary Studies, 7*(2), 76-83. http://www.journals.aiac.org.au/index.php/alls/article/view/2164

*Talan, T., & Batdi, V. (2020). Evaluating the flipped classroom model through the multi-complementary approach. *Turkish Online Journal of Distance Education, 21*(4), 31-67. https://doi.org/10.17718/tojde.803351

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research, 81*(1), 4-28. https://doi.org/10.3102/0034654310393361

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

Tamim, R. M., Borokhovski, E., Pickup, D., Bernard, R. M., & El Saadi, L. (2015). *Tablets for teaching and learning: A systematic review and meta-analysis.* Commonwealth of Learning. http://oasis.col.org/handle/11599/1012

*Tekbiyik, A., & Akdeniz, A. R. (2010). A meta-analytical investigation of the influence of computer assisted instruction on achievement in science. *Asia-Pacific Forum on Science Learning and Teaching, 11*(2). https://www.eduhk.hk/apfslt/v11_issue2/tekbiyik/index.htm

*Thompson, C. G., & von Gillern, S. (2020). Video-game based instruction for vocabulary acquisition with English language learners: A Bayesian meta-analysis. *Educational Research Review, 30*, 100332. https://doi.org/10.1016/j.edurev.2020.100332

*Timmerman, C. E., & Kruepke, K. A. (2006). Computer-assisted instruction, media richness, and college student performance. *Communication Education, 55*(1), 73-104. https://doi.org/10.1080/03634520500489666

*Tingir, S., Cavlazoglu, B., Caliskan, O., Koklu, O., & Intepe-Tingir, S. (2017). Effects of mobile devices on k-12 students' achievement: A meta-analysis. *Journal of Computer Assisted Learning, 33*(4), 355-369. https://doi.org/10.1111/jcal.12184

*Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning, 35*(3), 407-420. https://doi.org/10.1111/jcal.12347

*Torgerson, C. J., & Elbourne, D. (2002). A systematic review and meta-analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling. *Journal of Research in Reading, 25*(2), 129-143. https://doi.org/10.1111/1467-9817.00164

*Tsai, Y.-L., & Tsai, C.-C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education, 125*, 345-357. https://doi.org/10.1016/j.compedu.2018.06.020

*Tsai, Y.-L., & Tsai, C.-C. (2020). A meta-analysis of research on digital game-based science learning. *Journal of Computer Assisted Learning, 36*(3), 280-294. https://doi.org/10.1111/jcal.12430

Turban, E., Outland, J., King, D., Lee, J. K., Liang, T.-P., & Turban, D. C. (2018). *Electronic commerce 2018: A managerial and social networks perspective*. Springer. https://doi.org/10.1007/978-3-319-58715-8

US Department of Health, Education & Welfare. (1965). *The Vocational Education Act of 1963*. https://files.eric.ed.gov/fulltext/ED019402.pdf

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research, 85*(4), 475-511. https://doi.org/10.3102/0034654314564881

*Vo, H. M., Zhu, C., & Diep, N. A. (2017). The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation, 53*, 17-28. https://doi.org/10.1016/j.stueduc.2017.01.002

*Wandera, S. (2017). *Continuing the conversation about face-to-face, online, and blended learning a meta-analysis of empirical literature 2006-2017* (Publication No. 10637800) Wilmington University (Delaware). ProQuest Central; ProQuest Dissertations & Theses Global.

*Wang, C.-p., Lan, Y.-J., Tseng, W.-T., Lin, Y.-T. R., & Gupta, K. C.-L. (2020). On the effects of 3D virtual worlds in language learning — a meta-analysis. *Computer Assisted Language Learning, 33*(8), 891-915. https://doi.org/10.1080/09588221.2019.1598444

Wang, M., & Nunes, M. B. (2019). Matching serious games with museum's educational roles: Smart education in practice. *Interactive Technology and Smart Education, 16*(4), 319-342. https://doi.org/10.1108/ITSE-03-2019-0013

*Wenting, M., Adesope, O. O., Nesbit, J. C., & Qing, L. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901-918. https://doi.org/10.1037/a0037123

*Williams, S. L. (2006). The effectiveness of distance education in allied health science programs: A meta-analysis of outcomes. *American Journal of Distance Education, 20*(3), 127-141. https://doi.org/10.1207/s15389286ajde2003_2

*Wilson, A. B., Brown, K. M., Misch, J., Miller, C. H., Klein, B. A., Taylor, M. A., Goodwin, M., Boyle, E. K., Hoppe, C., & Lazarus, M. D. (2019). Breaking with tradition: A scoping meta-analysis analyzing the effects of student-centered learning and computer-aided instruction on student performance in anatomy. *Anatomical Sciences Education, 12*(1), 61-73. https://doi.org/10.1002/ase.1789

Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities? A meta-analysis. *Journal of Learning Disabilities, 51*(1), 73-84. https://doi.org/10.1177/0022219416688170

*Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology, 51*(6), 1991-2005. https://doi.org/10.1111/bjet.13023

*Xu, Z., Banerjee, M., Ramirez, G., Zhu, G., & Wijekumar, K. (2019). The effectiveness of educational technology applications on adult English language learners' writing quality: A meta-analysis. *Computer Assisted Language Learning, 32*(1-2), 132-162. https://doi.org/10.1080/09588221.2018.1501069

*Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension: A meta-analysis. *British Journal of Educational Technology, 50*(6), 3119-3137. https://doi.org/10.1111/bjet.12758

*Yammine, K., & Violato, C. (2015). A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. *Anatomical Sciences Education, 8*(6), 525-538. https://doi.org/10.1002/ase.1510

*Yun, J. (2011). The effects of hypertext glosses on l2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning, 24*(1), 39-58. https://doi.org/10.1080/09588221.2010.523285

Zhao, Y., Lei, J., Lai, B. Y. C. & Tan, H. S. (2005). What makes the difference? A practical analysis of research on the effectiveness of distance education. *Teachers College Record, 107*(8), 1836-1884. https://doi.org/10.1111/j.1467-9620.2005.00544.x

*Zheng, B., Warschauer, M., Lin, C.-H., & Chang, C. (2016). Learning in one-to-one laptop environments: A meta-analysis and research synthesis. *Review of Educational Research, 86*(4), 1052-1084. https://doi.org/10.3102/0034654316628645

Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review, 17*(2), 187-202. https://doi.org/10.1007/s12564-016-9426-9

*Zheng, L., Bhagat, K. K., Zhen, Y., & Zhang, X. (2020). The effectiveness of the flipped classroom on students' learning achievement and learning motivation: A meta-analysis. *Educational Technology & Society, 23*(1), 1-15. https://eric.ed.gov/?id=EJ1255782

*Zheng, L., Zhang, X., & Cui, P. (2020). The role of technology-facilitated peer assessment and supporting strategies: A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(3), 372-386. https://doi.org/10.1080/02602938.2019.1644603

*Zhu, G. (2021). Is flipping effective? A meta-analysis of the effect of flipped instruction on K-12 students' academic achievement. *Educational Technology, Research and Development, 69*(2), 733-761. https://doi.org/http://dx.doi.org/10.1007/s11423-021-09983-6

*Ziegler, N. (2016). Synchronous computer-mediated communication and interaction. *Studies in Second Language Acquisition, 38*(3), 553-586. https://doi.org/10.1017/S027226311500025X

Zucker, T. A., Moody, A. K., & McKenna, M. C. (2009). The effects of electronic books on pre-kindergarten-to-grade 5 students' literacy and language outcomes: A research synthesis. *Journal of Educational Computing Research, 40*(1), 47-87. https://doi.org/10.2190/EC.40.1.c

Appendix A: Overview of the Methodology

Second-order meta-analysis (or meta-analysis of meta-analyses), to be implemented here, belongs to a broader category of systematic reviews of research literature and employs the same basic methodology (e.g., Cooper, 2017) as any first-level (or "primary") meta-analytical investigation.

"Systematic review" is a general term that encompasses both quantitative and qualitative methodologies that are appropriately targeted, inclusive, transparent, reproducible, methodologically sound and rigorously interpreted (Bethel & Bernard, 2010; Cooper & Hedges, 1994). Although it is perhaps the best known, meta-analysis is not the only quantitative synthesis methodology that is sometimes employed to characterise a population of studies around a set of given research questions. Narrative and vote-count methodologies, for example, can also be used, but offer less precision in terms of quantitative analyses and inferences.

Among systematic reviews, meta-analysis is a specific class of research syntheses that relies on quantitative data from a multitude of primary studies addressing a common core research question or a set of closely related and/or complementary research questions.

Meta-analysis summarises systematically collected effect sizes from individual studies to estimate either the magnitude of a difference between groups of interest (*d*-family effect size) or degree of association between variables of interest (*r*-family effect size) in the entire related population. Furthermore, meta-analyses aim at exploring and explaining the variability that surrounds the overall effect size by systematically coding and analysing methodological, substantive, contextual and demographic moderator variables. The main research question should be stated, substantiated and operationalised a priori to inform search strategies, set up and describe inclusion criteria, and meaningfully guide the review process through all its steps (e.g., as outlined by Cooper, 2017) from study selection, through effect size extraction, aggregation and analyses towards interpretation and presentation of the findings.

Meta-analysis as Quantitative Synthesis

The basic metric and the unit of analysis in meta-analytical research is an effect size. Most frequently used in comparative studies in the social sciences in general, and especially in education, is the *d*-type effect size, which is simply the standardised difference between the means of two groups that are compared to each other on a consistent set of relevant outcomes. Basic equations used in typical comparative meta-analyses are shown below.

The *d*-type effect size (also known as Cohen's *d*) is the standardised mean difference of two groups. Usually these are the treatment condition — that is, the one that implements the intervention whose effectiveness is to be assessed — and the control condition — that is, the "reference-point" or "business-as-usual" intervention to which the experimental one is compared. In non-interventional studies these groups could represent categories of participants — for example, gender or geographical location. The respective means — $\bar{X}_E$ and X_C— are subtracted from one another and standardised, as depicted in Equation 1 (below):

Equation 1

$$d = \frac{\bar{X}_E - \bar{X}_C}{SD_{Pooled}}$$

The denominator is the pooled standard deviation of both groups, or the standardisation term, calculated according to Equation 2:

Equation 2

$$SD_{Pooled} = \sqrt{\frac{(n_E - 1)SD^2_E + (n_C - 1)SD^2_C}{(n_E - 1) + (n_C - 1)}}$$

An effect size of *d*-type can also be calculated or estimated from various inferential statistics, such as *t*-tests, *F*-ratios, associated *p*-values (exact or in relation with the pre-set alpha levels) or a combination using an array of formulas (see, for example, Bernard, Borokhovski, Schmid, Tamim & Abrami, 2014; Glass et al., 1981; Hedges et al., 1989). Regardless of what specific formula is used, the sample size of both groups (or at least a total number of participants to be arbitrarily split in most likely proportions) must be available, and for non-signed statistics, the direction of the effect must be indicated in the report to enable reasonably reliable effect size extraction.

There is also a correction to Cohen's *d* for potential bias associated with small samples, known as Hedges' *g*. Equation 3 is used to calculate it:

Equation 3

$$g \cong d\left[1 - \frac{3}{4N - 9}\right]$$

Effect sizes from individual primary studies are then summarised to reflect the overall effectiveness of the treatment in question across empirical research on the topic as follows.

Effect Sizes' Aggregation

To produce a truly representative overall average, effect sizes are weighted at the synthesis phase of a meta-analysis according to one of two analytical models: the fixed effect model or the random effects model. Comprehensive descriptions of the conceptual underpinnings and procedural details of these two models can be found in the literature on meta-analysis (e.g., Borenstein et al., 2009, 2010; Hedges

& Olkin, 1985; Pigott, 2012). Typically, for meta-analyses in the social sciences, particularly in education, the random effects model is deemed to be more suitable when there is a great variety of samples, settings and contexts in the reviewed research. With respect to a second-order meta-analysis, all the same logic applies with a few adjustments that we need to keep in mind and are presented in the following section.

Second-Order Meta-analyses

With the increasing number of published meta-analyses in a variety of social science areas, a second-order meta-analysis allows for a more systematic and reliable methodology for synthesising related results than a narrative review. Moreover, because it enables the synthesis of effect sizes from different meta-analyses while considering the standard errors, it is more adequate than vote counts. Using such a methodology, researchers can benefit from published literature while reaching more generalisable findings than individual studies or regular meta-analyses can offer. This is particularly true regarding effect size because of the larger included sample size. As such, a second-order meta-analysis is best suited for answering big questions pertaining to a particular area of research with a considerable number of publicly available meta-analyses without the need to replicate their findings by running a huge new meta-analysis.

Because of the nature of the literature it is aiming to synthesise, a meta-analysis is limited by the amount of information provided in the documents under review and the quality of the reports themselves. With a second-order meta-analysis, this issue is magnified; the review is further limited by the information presented in each meta-analysis, which is farther from the original data collected by the original empirical research. This is somewhat expected when one is working with a review of reviews; after all, we are trying to synthesise a set of syntheses that should by nature be succinct and condensed regarding certain aspects pertaining to the original primary studies. Furthermore, and because of the inability to extract and code for a variety of contextual features, it is usually difficult to conduct comprehensive moderator analyses.

It is important to note that the second-order meta-analysis approach used here was piloted, tested and validated previously with a substantive body of meta-analyses addressing technology integration in face-to-face educational contexts (Tamim et al., 2001). The validation process and its findings (that agreed with the results from the second-order meta-analysis) proved that the methodology is an adequate and reliable technique for synthesising effect sizes and estimating the average effect size in relation to a specific phenomenon.

Specific aspects of the implementation (e.g., search strategy, inclusion criteria, decision-making justification, analytical procedures employed, etc.) are outlined in greater detail in the Methodology section.

Appendix B: Second-Order Meta-analysis of Education Technology Applications: Search History

Overview

Source	Initial Results	After Duplicates	Post-2000
ERIC	354	353	295
Education Source	360	265	245
ProQuest Education	199	101	99
Branching	2	2	2
Total:	915	721	641

ERIC (EBSCO)

Search performed October 14, 2021

(DE "Instructional Systems" OR DE "Audience Response Systems" OR DE "Audiovisual Aids" OR DE "Audiovisual Communications" OR DE "Audiovisual Instruction" OR DE "Influence of Technology" OR DE "Technology Integration" OR DE "Technology Uses in Education" OR DE "Electronic Equipment" OR DE "Educational Media" OR DE "Multimedia Instruction" OR DE "Multimedia Materials" OR DE "Electronic Journals" OR DE "Computer Managed Instruction" OR DE "Integrated Learning Systems" OR DE "Information Technology" OR DE "Educational Technology" OR DE "Internet" OR DE "Video Technology" OR DE "Blended Learning" OR DE "Electronic Learning" OR DE "Handheld Devices" OR DE "Laptop Computers" OR DE "Computer Mediated Communication" OR DE "Computer Peripherals" OR DE "Computer Simulation" OR DE "Computer Software" OR DE "Social Media" OR DE "Computer Assisted Instruction" OR DE "Computer Use" OR DE "Computer Uses in Education" OR DE "Computers" OR DE "Distance Education")

AND

(effective* OR perform* OR achieve* OR success* OR GPA OR grades)

AND

(AB meta-analysis OR TI meta-analysis OR DE meta-analysis)

Limiters - Publication Type: Reports - Descriptive, Reports - Evaluative, Reports - Research, Reports - Research-practitioner Partnerships; Language: English

RESULTS: 354 (353 after duplicates)

NOTE: the date limit of 2000 was left out; after pre-2000 items were removed there were **295** publications left for abstract screening.

Education Source (EBSCO)

Search performed October 14, 2021

DE “Student response systems” OR DE “Calculators in education” OR DE “Mobile learning” OR DE “Virtual classrooms” OR DE “Virtual reality in education” OR DE “Digital learning” OR DE “Instructional systems” OR DE “Integrated learning systems” OR DE “Audience response” OR DE “Audiovisual aids in education” OR DE “Audiovisual education” OR DE “Educational technology” OR DE “Digital media” OR DE “Digital video” OR DE “Digital communications” OR DE “Multimedia systems in education” OR DE “Computers in education” OR DE “Computer simulation” OR DE “Computer software” OR DE “Information technology” OR DE “Computer networks” OR DE “Computers” OR DE “Video games” OR DE “Video games in education” OR DE “Distance education” OR DE “Television in education” OR DE “Blended learning” OR DE “Computer assisted instruction” OR DE “Courseware” OR DE “Cyberschools” OR DE “Digital badges in education” OR DE “Internet in education” OR DE “Online education” OR DE “Open universities” OR DE “Telecommunication in education” OR DE “Telecourses” OR DE “Virtual schools” OR DE “Virtual universities & colleges”

AND

AB meta-analysis OR TI meta-analysis OR meta-analysis

AND

effective* OR perform* OR achieve* OR success* OR GPA OR grades

Limiters: English

Filtered: Magazines (19), Books (1)

RESULTS: 360 (265 after duplicates)

NOTE: the date limit of 2000 was left out; after pre-2000 items were removed there were **245** publications left for abstract screening.

ProQuest Education Database

Search performed October 13, 2021

((MAINSUBJECT.EXACT(“Computer assisted instruction CAI”) OR AINSUBJECT.EXACT(“Computers”) OR MAINSUBJECT.EXACT(“Audiovisual communications”) OR MAINSUBJECT.EXACT(“Distance learning”) OR MAINSUBJECT.EXACT(“Digital electronics”) OR MAINSUBJECT.EXACT(“Interactive computer systems”) OR MAINSUBJECT.EXACT(“Educational software”) OR MAINSUBJECT.EXACT(“Interactive learning”) OR MAINSUBJECT.EXACT(“Multimedia computer applications”) OR MAINSUBJECT.EXACT(“Digital media”) OR MAINSUBJECT.EXACT(“Computer use”) OR MAINSUBJECT.EXACT(“Artificial intelligence”) OR MAINSUBJECT.EXACT(“Digital technology”) OR MAINSUBJECT.EXACT(“Multimedia”) OR MAINSUBJECT.EXACT(“Video teleconferencing”) OR MAINSUBJECT.EXACT(“Handheld computers”) OR MAINSUBJECT.EXACT(“Educational technology") OR MAINSUBJECT.EXACT("Audiovisual materials") OR MAINSUBJECT.EXACT("Multimedia communications") OR MAINSUBJECT.EXACT("Software") OR MAINSUBJECT.EXACT("Blended learning") OR MAINSUBJECT.EXACT("Video equipment") OR MAINSUBJECT.EXACT("Programmed instruction") OR MAINSUBJECT.EXACT("Online instruction") OR MAINSUBJECT.EXACT("Computer & video games") OR MAINSUBJECT.EXACT("Internet") OR MAINSUBJECT.EXACT("Mobile content") OR MAINSUBJECT.EXACT("Online tutorials") OR MAINSUBJECT.EXACT("Simulators") OR MAINSUBJECT.EXACT("Digital video") OR MAINSUBJECT.EXACT("Digital computers") OR MAINSUBJECT.EXACT("Flipped classroom")) AND noft(effective* OR perform* OR achieve* OR success* OR GPA OR grades)) AND (ab(meta-analysis) OR ti(meta-analysis) OR su(meta-analysis))

Limited by: Source type:5 types searched Conference Papers & Proceedings, Dissertations & Theses, Reports, Scholarly Journals, Working Papers Language:English

RESULTS: 199 (101 after duplicates)

NOTE: the date limit of 2000 was left out; after pre-2000 items were removed there were **99** publications left for abstract screening.

Branching

Performed December 29, 2021

Several articles listing previous meta-analyses on educational technology were identified and we performed citation searching to ensure we had them all in our collection.

The articles branched for additional citations were:

Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education, 105*, 14-30. https://doi.org/10.1016/j.compedu.2016.11.005

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

Xie, C., Cheung, A. C. K., Lau, W. W. F., & Slavin, R. E. (2020). The Effects of Computer-Assisted Instruction on Mathematics Achievement in Mainland China: A Meta-Analysis. *International Journal of Educational Research, 102*, 101565. https://doi.org/10.1016/j.ijer.2020.101565

After reviewing these publications, an additional two studies were included for abstract screening.

LEARNING FOR SUSTAINABLE DEVELOPMENT

COL is an intergovernmental organisation created by Commonwealth Heads of Government to encourage the development and sharing of open learning and distance education knowledge, resources and technologies.

VISION: To be the foremost global agency that promotes learning for sustainable development.

MISSION: To help governments and institutions to expand the scale, efficiency and quality of learning by using open, distance and technology-based approaches.

COMMONWEALTH OF LEARNING

4710 Kingsway, Suite 2500
Burnaby, BC V5H 4M2 Canada
Phone + 1 604 775 8200 / Fax + 1 604 775 8210
info@col.org
www.col.org
@col4d
facebook.com/COL4D/
instagram.com/commonwealthoflearning/

August 2022