Computers & Education 133 (2019) 27–42

Contents lists available at ScienceDirect

Computers & Education

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

How is the use of technology in education evaluated? A systematic review

Jennifer W.M. Lai*, Matt Bower

Department of Educational Studies, Macquarie University, Sydney, NSW, 2109, Australia

ARTICLE INFO

Keywords:
Evaluation methodologies
Interactive learning environments
Applications in subject areas

ABSTRACT

There are a large variety of methodologies, contexts and perspectives that have been used to evaluate the use of technology in education. The vast array of literature involving learning technology evaluation makes it challenging to acquire an accurate sense of the different aspects of learning that are evaluated, and the possible approaches that can be used to evaluate them. This study conducted a systematic review of how technology use in education has been evaluated, based on 365 papers published in Computers and Education between 2015 and 2017. The papers analyzed encompassed a diverse range of education levels, disciplines and technologies, that were broadly commensurate with characterizations of the learning technology field from other studies. The analysis found that the evaluation of learning technology use tended to focus on eight themes: learning outcomes, affective elements, behaviors, design, technology elements, pedagogy, presence, and institutional environment. In the majority of studies (66.6%) at least one established instrument was used to evaluate the effectiveness of learning technology usage, however, a wide variety of instruments were observed within sub-themes of evaluation. The 22 instruments used to evaluate an aspect of learning technology usage in more than one study are identified. Evaluation trends for different disciplines, technologies and educational levels are also established. The findings provide an analytical framework that educators and researchers can utilize when evaluating the use of technology in education, and could potentially inform the development of new, more holistic and more robust evaluation methods.

1. Background and motivation for this study

With the continual influx of new and emerging technologies available for use in education, it is critical to evaluate the degree to which learning technology usage contributes to learning and teaching (Iriti, Bickel, Schunn, & Stein, 2016). There are different motivations for integrating technology into learning, for instance, to improve student learning outcomes, improve access to learning, and enhance learner motivation (Bower, 2017). Many high quality meta-analyses show consistent, moderately positive effect sizes for the use of technology (Clark, Tanner-Smith, & Killingsworth, 2016; Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014), yet, the perceived educational effectiveness of using learning technology is very much determined by how it is evaluated (Phillips, McNaught, & Kennedy, 2012).

In general, evaluation can be characterized as "the process by which people make judgements about value and worth" (Oliver, 2000, p. 20). Researchers frequently evaluate the application of learning technologies to better understand how technology supports learners (e.g., Bulu, 2012; Claros, Cobos, & Collazos, 2016; Conole, De Laat, Dillon, & Darby, 2008; Foshee, Elliott, & Atkinson, 2016;

* Corresponding author.
E-mail addresses: jennifer.lai@mq.edu.au (J.W.M. Lai), matt.bower@mq.edu.au (M. Bower).

https://doi.org/10.1016/j.compedu.2019.01.010
Received 15 May 2018; Received in revised form 16 January 2019; Accepted 20 January 2019
Available online 21 January 2019
0360-1315/

Pegrum, Howitt, & Striepe, 2013; Wang, Shannon, & Ross, 2013, etc.). However, in the context of learning technology, this judgement process is complex and often controversial. There is a plethora of instruments, orientations and aspects that researchers may use to evaluate learning technology implementations, spread across hundreds if not thousands of studies. This can make it difficult for researchers and teachers to determine how they should evaluate the educational use of technologies.

Recent reviews relating to the use of learning technology typically relate to one specific learning technology like mobile learning (e.g., Crompton, Burke, & Gregory, 2017; Hwang & Tsai, 2011; Liu et al., 2014; Wong & Looi, 2011; Wu et al., 2012); e-Portfolios (Beckers, Dolmans, & van Merriënboer, 2016); e-learning, online learning or MOOCs (e.g., Kennedy, 2014; Means, Toyama, Murphy, Bakia, & Jones, 2010); microblogging and social media (e.g., Gao, Luo, & Zhang, 2012; Manca & Ranieri, 2013; Tess, 2013); computer games, digital games or serious games (e.g., Boyle et al., 2016; Calderón & Ruiz, 2015; Cheng, Chen, Chu, & Chen, 2015; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Kordaki & Gousiou, 2017; Petri & Gresse von Wangenheim, 2017; Young et al., 2012); and augmented reality (AR) or virtual environments (e.g., Bacca, Baldiris, Fabregat, Graf, & Kinshuk, 2014; Hew & Cheung, 2010; Mikropoulos & Natsis, 2011). Other studies focus on the impact of learning technology usage on a particular type of learner, for instance, elementary students (e.g., Chauhan, 2017); primary school students (e.g., Hainey, Connolly, Boyle, Wilson, & Razak, 2016); higher education students (e.g., Bernard, Borokhovski, Schmid, Tamim, & Abrami, 2014; Kirkwood & Price, 2014), or pre-service teachers (e.g., Baran, 2014). Yet other studies have concentrated on the influence of learning technology usage on a single aspect of learning like learning effectiveness or learning outcomes (e.g., Cheung & Slavin, 2012; Mohsen, 2016; Sung, Chang, & Liu, 2016) or learning skills (e.g., de Klerk, Veldkamp, & Eggen, 2015; Wollscheid, Sjaastad, & Tømte, 2016). Some studies only focus on single discipline, for instance, science (e.g., Smetana & Bell, 2012).

However, at the time of writing, the researchers (who examined several hundred papers as part of their investigation) were only able to source two review papers that primarily focused on how the use of technology is *evaluated* in the field of education. In their review of how the quality of serious games were evaluated, Calderón and Ruiz (2015) identified characteristics such as satisfaction, engagement, usefulness, usability and performance as the focus of evaluation. In a systematic review of how games for computing education are evaluated, Petri and Gresse von Wangenheim (2017), identified 43 different factors within different aspects including learning effect, motivation, user experience, usability and instructional aspects. These two review studies shed some initial light on how learning technology use is evaluated, but they relate to the very specific technological context of educational games, meaning that it is not possible to gauge whether the aspects and approaches identified apply for other technologies.

Six other review papers which focused on determining educational effects (rather than how technology was evaluated) did examine up to four different aspects in their studies (Beckers et al., 2016; Cheung & Hew, 2009; Cook et al., 2012; Means et al., 2010; Pérez-Sanagustín et al., 2017; Salleh, Mendes, & Grundy, 2011). All six papers assessed learning outcomes as well as affective elements of learning. The review by Cook et al. (2012) evaluated teaching strategies, whereas Salleh et al. (2011) investigated the design of courses in their review. However, these reviews do not capture or detail all of the different educational aspects of learning technology use that were evaluated across the studies, nor systematically review the instruments that were used, nor examine evaluation patterns across various technologies, disciplines and education levels. In addition, each literature review adopts different categorizations, sample sizes, selection criteria and so on, meaning that it is not possible to compare or combine results from the various studies on a consistent basis in a way that would accurately characterize education technology evaluation.

In this study we conducted a systematic review of the various approaches to evaluation that have been used within the empirical educational technology literature, to provide educators and researchers with a more accurate sense of evaluation practices so that they may conduct more effective educational technology evaluation. The overarching research question for this study is "*How are educational uses of technology evaluated?*". To focus the study, this overarching question is divided into the following four sub-questions:

RQ1 What constructs are examined when evaluating the use of technology in education?
RQ2 What research methods are used in studies that evaluate the use of technology in education?
RQ3 What are the prevailing instruments that are used to evaluate the use of technology in education?
RQ4 What is the relationship between the educational aspects being evaluated and the technologies, disciplines and levels of education being investigated?

2. Methodology

The methodological approach used in this paper was a systematic literature review. Systematic literature reviews play a vital role both in supporting further research effort and providing an unbiased synthesis and interpretation of the findings in a balanced manner (Kitchenham et al., 2010). This type of review attempts to collate relevant evidence that fits the pre-specified eligibility criteria to answer specific research questions (Moher, Liberati, Tetzlaff, & Altman, 2009, p. 3). The review process was divided into three steps proposed by Kitchenham and Charters (2007), i.e. planning, conducting and reporting the systematic review. Reporting of the review was guided by the principles of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Moher et al., 2009). In particular, the PRISMA methodology was used to guide the description of the eligibility criteria, information sources, data collection process, data items, and synthesis of results.

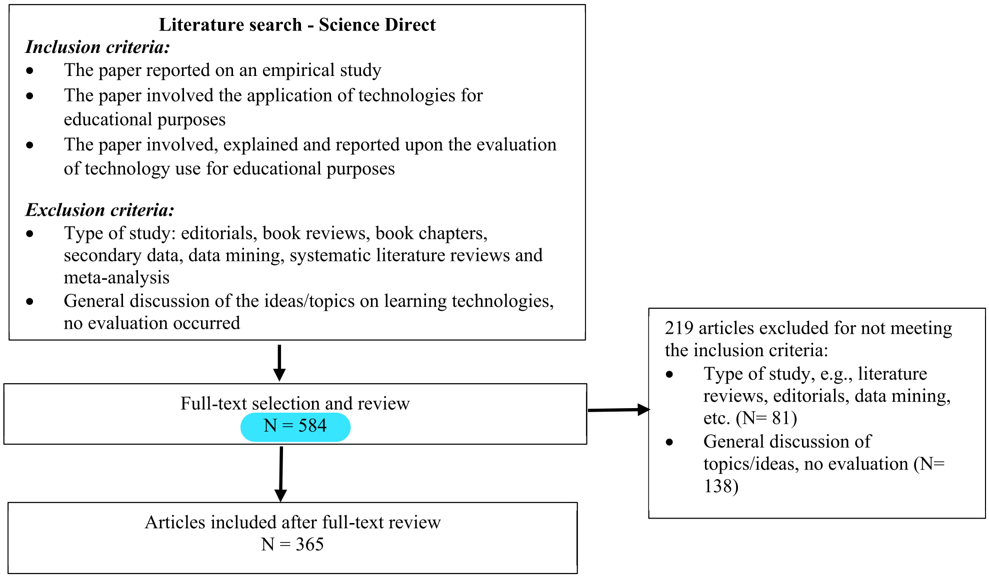

Fig. 1. A diagrammatic representation of the literature selection process.

2.1. Planning of the review

2.1.1. Paper selection

Broadly characterizing the evaluation of learning technology usage in a way that is systematic and deep is inherently challenging, because of the large number of evaluation studies that have been published and the substantial work involved in conducting systematic reviews. The decision was made to select a large and inclusive but also recent sample of papers from a high quality journal in order for the review to a) encapsulate learning technology evaluation across a wide variety of contexts so that themes and patterns could be detected, b) have some assurance that evaluation practices were contemporary as well of high quality. Accordingly, studies were selected from Computers & Education research articles published between January 2015 and December 2017. Computers & Education has the highest impact factor of any learning technology journal that spans all technologies, subject areas, and levels of education (2017 impact factor 4.538, 5-year impact factor of 5.568) (Clarivate Analytics, 2019). This journal is ranked the top journal of "Educational Technology" and "Social Science" categories in Google Scholar, with a h5-index of 91 and a h5-median of 152 (Google, 2019). Thus, it could be relied upon for broad and high quality empirical research relating to the evaluation of technology usage in education across a wide variety of contexts. Constraining the review to papers published between 2015 and 2017, the previous three years at the time of conducting the review, meant that contemporary evaluation practices could be examined, and that the review was logistically feasible given the hundreds of papers examined.

The approach to paper selection is shown in Fig. 1. There were a total of 584 articles published in Computers & Education between January 2015 (volume 80) and December 2017 (volume 115). Due to the nature of this systematic literature review, the researchers initially conducted a review all of the papers, i.e. 584 papers in the original sample. The papers were sourced from *Computers and Education* series pages appearing in the Science Direct database. The abstracts, introductions, methodology, findings and conclusions were examined, and the decision about whether to include the paper was based upon whether it met the following criteria:

1 The paper reported on an empirical study
2 The paper involved the application of technologies for educational purposes
3 The paper involved, explained and reported upon the evaluation of technology use for educational purposes

These widely inclusive selection criteria resulting in a large sample size enabled broad characterization of approaches to the evaluation of learning technology usage, as well as potentially allowing more valid generalization of results. The inclusive sampling process also provided greater opportunity for recognizing patterns and differences across and between contexts.

According to the selection criteria, editorials, book reviews, papers involving secondary data or data mining, literature reviews, meta-analyses, and general ideas on technologies (e.g., Facebook) were not included because they did not involve the empirical evaluation of technology use in education. If there was any ambiguity about whether to include a paper in the review, the authorial team discussed the paper until a consensus was reached. This resulted in the selection of 365 papers for inclusion in the study (see Table 1). This represents a much larger sample size than typical scholarly reviews in the educational technology field. For instance, of the 32 reviews cited in the introduction, the average number of papers included in the sample was 73 papers, with a range of 10–176 papers. A Z-test (Rea & Parker, 2014) indicated that the slight drop in the percentage of empirical articles over the three-year period shown in Table 1 was not statistically significant ($Z = -0.885$, $p = 0.376$).

Table 1
Paper selection results.

	2015	2016	2017	Total
Number of articles examined	250	172	162	584
Number of articles selected for this study	162	105	98	365
Percentages of number of articles selected	64.8%	61.0%	60.5%	62.5%

2.2. Conducting the review

2.2.1. Approach to analysis

Each of the 365 papers was read in full to determine the context and evaluation methods used. Relevant information from each selected paper was systematically extracted in accordance with the defined research questions, with an Excel spreadsheet being used to store and analyze all data. The approach to harvesting data from the papers and analyzing them can be described as occurring in two broad phases, noting that the analysis approach did at times involve a degree of iteration during thematic analysis to refine themes and category boundaries. The two phases were as follows.

2.2.1.1. Phase 1. The Excel spreadsheet columns were initially divided into three sections, where the first section recorded the basic information of the reviewed papers. Columns were included for the *Year of publication(YP); Unique number that we allocated to each journal article for classification purpose(NC); Title of the paper(TP); Volumes and Issues(VI); Authors' details(AU); Abstract(AB); Status, i.e. accepted or rejected for review and the reasons why(ST)*. The second section of the spreadsheet recorded contextual aspects of the study, with columns for the *Levels of education(LE), Discipline areas being taught(DA) and Technologies used(TU)*. This served to broadly characterize the types of technology-enhanced learning evaluations being conducted.

The third part of the spreadsheet addressed the research questions. In order to record the characteristics of themes/aspects of the reviewed studies and addressed RQ1, we included the column, i.e. *What was being evaluated (WH),* e.g. *perceived ease of use,* or *frequency of use, etc*. If a paper examined more than one construct of learning technology usage (which many papers did), we added additional columns to make sure that we captured all the constructs evaluated in the study. Columns about *Research methodology(RM)* and *Types of data(TD)* were used to address RQ2. The research methods used in each paper (RQ2) were categorized according to how the authors of each paper *described* their method, for instance, experiments, quasi-experiments, surveys, case studies, and so on. Previous systematic literature reviews, such as that by Crompton et al. (2017), have used similar categorizations to describe research approaches. Additionally, the nature of the actual data used to conduct each evaluation was recorded, as to whether it was purely quantitative (including Likert scales), qualitative, or both (i.e. mixed). For RQ3, we included columns for *How the constructs were evaluated(HO)*, e.g. quantitative data with Likert scale, and the *Details of the instruments(IN)* for each construct being evaluated.

2.2.1.2. Phase 2. After filling in the columns for all 365 papers, additional columns presenting the thematic analysis were added. For instance, different disciplines like physics, biology, chemistry were entered in the *Discipline areas being taught (DA)* column and Sciences was recorded in the next column *Discipline areas being taught –Theme(DAT)*. Table 2 shows the final organization of the analysis spreadsheet, including theme columns for *Levels of education - Theme (LET)*, *Discipline areas being taught –Theme(DAT)*, *Technologies used – Theme (TUT)*, *Research methodology – Theme (RMF)*, *Types of data –Theme (TDT)*, and *How the aspects were evaluated–Theme(HOT)*.

Themes and sub-themes were developed using general thematic analysis principles (Boyatzis, 1998). Many of the themes were relatively straight-forward to derive directly from the Phase 1 data, for instance, the methodologies used (RQ2) as well as the levels of education and the discipline areas. For the investigation of other areas, such as the constructs that were evaluated during educational technology research (RQ1), more extensive and iterative thematic analysis was required. This typically involved several discussions among the researcher team, based on repeated analysis of the primary data (articles), and several rounds of refinement, before final

Table 2
Details of the analysis spreadsheet.

Contents	Columns in the analysis spreadsheet
Basic information of the papers	Year of publication(YP); Unique number that we allocated to each journal article for classification purpose(NC); Title of the paper(TP); Volumes and Issues(VI); Authors' details(AU); Abstract(AB); Status(S), i.e. accepted or rejected and the reasons why(ST)
Contextual aspects of the study	Levels of education(LE), Levels of education –Theme(LET); Discipline Areas being taught (DA), Discipline areas being taught –Theme(DAT), Technologies used (TU), Technologies used –Theme(TUT)
RQ2: What research methods are used in the reviewed papers to evaluate the use of technology in education?	Research methodology(RM), Research methodology – Theme (RMT), Types of data (TD), Types of data –Theme (TDT)
RQ1. What constructs are examined when evaluating the use of technology in education?	What constructs were being evaluated (WH), What constructs were being evaluated–Theme(WHT)
RQ3. What are the prevailing instruments that are used to evaluate the use of technology in education?	How the themes/aspects were evaluated(HO), How the themes/aspects were evaluated–Theme(HOT), instruments(IN)

themes were decided. Thematic analysis was both a deductive and inductive process, both relying on starting points from other papers and researcher knowledge, but also using the data to authentically inform the creation of themes. In cases where theme boundaries were ambiguous the research team returned to the actual papers rather than simply relying on initial recorded categorizations, in order to uphold the accuracy of thematic classification. From each discussion, changes were recorded by adding or adjusting themes to updated versions of the spreadsheet, and returning to examine specific papers that were on theme boundaries if necessary. For instance, the construct 'emotional problems' was added to the 'anxiety or boredom' sub-theme in the second round of the analysis. 'Teaching quality/credibility' was combined with 'pedagogical practice, teaching strategies' and formed the overall 'Teaching/Pedagogy' theme. Having the research team work together to determine categorizations of ambiguous articles meant that 100% consensus was reached in each case. Because the data being analyzed could be objectively categorized once the categorization scheme was defined, researcher time was used to improve the quality of the final categorizations rather than calculating inter-rater reliability measures, the latter of which do not improve the quality of categorization across the entire dataset. Previous systematic literature reviews, for instance by de Klerk et al. (2015) and Petri and Gresse von Wangenheim (2017), also adopted this approach.

2.3. Reporting the review

Reporting drew upon the data extraction from Phase 1 and the thematic analysis from Phase 2 in order to summarize the contexts being investigated across the data corpus and address the research questions. Pivot tables were used to directly extract the education levels, disciplines and technologies being investigated amongst the various studies. A two-level pivot table based on spreadsheet columns *WH* and *WHT* was able to summarize the constructs being evaluated, categorized according to their various themes or 'aspects' (RQ1). The research methods including types of data were able to be directly extracted using single level pivot tables from columns *RMT* and *TDT*, respectively (RQ2). The research instruments that were used could be directly extracted from the respective spreadsheet column *IN* (RQ3).

To identify trends in the way that learning technology use is evaluated across different contexts (according to disciplines, levels, and technologies, i.e. RQ4), Z-tests were used to identify combinations of learning technology aspects and specific contexts that were more frequent or less frequent than indicated by the overall corpus (365 papers). The *Z*-tests were computed by dividing the difference between the observed and expected values by the standard deviation of this difference (Bakeman & Gottman, 1997). For instance, there was a significantly higher proportion of online discussion platform studies that investigated learning behaviors (such as interactivity between participants) than for other technologies. The individual studies were then examined in order to help explain the emergent patterns and trends. The researchers then conferred about the trends and the reasons for them, based on the paper and other literature, in order to infer the rationale for and validity of the claims.

To report the findings, the first sections of the Results characterize the evaluation contexts, including the levels of education, the discipline areas examined, and the technologies that were used. This is followed by reporting of findings according to the research questions, regarding the aspects of learning technology usage that were evaluated, research methods used, the instruments that were used, and trends that were observed. Critical interpretations of the results are reserved for the Discussion and Conclusion sections of the paper.

3. Results

3.1. Contexts being evaluated

The sample of 365 studies captured a wide range of learning technology evaluation contexts. Almost half (48.9%) of the studies related to university or college level education, more than a third relating to school education (either secondary or primary), with the remaining studies involving adult education, early childhood education, or studies across levels (see Table 3). These proportions broadly align with the previous studies (e.g., Kirkwood & Price, 2014), with the high proportion of tertiary studies potentially due to the expedience and availability of university samples.

Discipline areas represented in the sample were broadly classified in accordance with Hew and Cheung (2010), and are shown in Table 4. A wide range of discipline areas were represented amongst the sample, most prevalently science-oriented subjects (27.9%) and this result was similar to previous studies (Chauhan, 2017; Lee, Linn, Varma, & Liu, 2010). The high proportion of science studies

Table 3 Convenient sampling
Levels of education of the reviewed papers.

Levels	No. of papers	%
University/College	177	48.9%
Secondary school	76	20.9%
Primary school	62	17.0%
Adult learning/Open Education	25	6.9%
Mixed (e.g., Primary and secondary, Secondary and University)	13	3.6%
Early-childhood and Kindergarten	12	3.3%
Total	**365**	**100.0%**

Economic, political, academic funding leaverage towards STEM

Still imbalanced

Table 4
Discipline areas being taught of the reviewed papers.

Discipline areas being taught	No. of papers	%
Sciences (e.g., mathematics, physics, biology, chemistry or engineering)	102	27.9%
Information technology (e.g., information system or information and communications technology)	50	13.7%
Professional education and training (e.g. Teacher education/training programs, professional development)	41	11.2%
Multi-disciplines (e.g., sciences + literature + history + geography, sociology or psychology + humanities + music + information technology + business + mathematics + chemistry)	40	11.0%
Language and Literacy (e.g., English, Chinese, Italian, English as a foreign language or Chinese as a foreign language)	39	10.7%
Business and Economics (e.g., marketing, tourism, or accounting)	20	5.5%
Medical-related programs/courses (e.g., nursing, first aid, medical school or health education)	19	5.2%
Arts (e.g., art and design, performing arts – dance or music)	18	4.9%
General development of skills and knowledge (e.g., problem-solving skills or presentation skills)	15	4.1%
Psychology, Sociology or Humanities	15	4.1%
Geography or History	6	1.6%
Total	**365**	**100.0%**

reflected in the technology-enhanced learning literature may be due to the way in which technology is suited to supporting scientific enquiry (Looi et al., 2011). The fact that technology is the subject of the information technology discipline means that it is often used to support learning, accounting for the high proportion of studies in this area (Harris, Mishra, & Koehler, 2009).

There was also a high proportion of studies that focused on professional education and training, for instance, how technology could be effectively used for teachers to share professional knowledge (Ekanayake & Wishart, 2015). Using technology to teach language and literacy was also very common (Chun, Kern, & Smith, 2016; Prior, Mazanov, Meacheam, Heaslip, & Hanson, 2016). Other discipline areas with lesser representation included business and economics, medically-related studies, arts, related to the areas of general skills development, psychology, sociology or humanities, geography and history. There was also a considerable proportion of studies that involved a mixed of disciplines (as outlined in Table 4).

There was also a diverse range of technologies that were being studied, with the most popular technologies being games (18.4%), Web 2.0 tools (14.8%), and mobile learning (14.8%) (see Table 5). The results were aligned with previous reviews, for instance that there has been a surge of interest in the use of games and mobile devices for educational purposes (Cheng, Lin, & She, 2015; Wu et al., 2012). Other technologies included those relating to digital instruction, online learning management systems, animations and simulations, feedback systems, Massive Open Online Courses (MOOCs), student response systems, programming technologies, augmented reality, robotics, and e-books. The identification of so many different types of learning technologies highlights the diversity of the educational technology ecology (in accordance with Bower, 2016).

The types and proportions of educational levels, discipline areas and technologies not only describe the contexts in which the use of technology in education was being evaluated, but also provides an interesting characterization of recent empirical learning technology research more broadly.

Table 5
Technologies used in the reviewed papers.

Technologies	No. of papers	%
Games/Mobile games in different disciplines	67	18.4%
Web 2.0 learning technologies (e.g., social media, Social Network systems, Wiki or Blogs)	54	14.8%
Mobile learning (e.g., tablets, iPads, computers, interactive tools/technologies or mobile device)	54	14.8%
Virtual world/virtual reality	24	6.6%
Digital instructions or instructional visual aids	22	6.0%
Management systems (e.g., classroom management systems, learning management systems or self-regulated learning systems)	21	5.8%
Animations and simulations (e.g., instructional animation or computer animation)	18	4.9%
Discussion/Online discussion platforms (e.g., online interaction platform, online collaborative network or collaborative simulation)	18	4.9%
Online learning course delivery, e-learning or non-MOOC	16	4.4%
Blended learning (i.e. use of technology with face-to-face learning)*	12	3.3%
Technology-enhanced feedback system, online feedback system or audio feedback system	11	3.0%
MOOCs	10	2.7%
Student response system (SRS)	10	2.7%
Programming	8	2.2%
Embodied agents, non-player agents, pedagogical agents or teachable agents	7	1.9%
Augmented reality (AR) technology	5	1.4%
Robotics	4	1.1%
Online book, e-books or digital storytelling	4	1.1%
Total	**365**	**100.00%**

Note: * Blended learning system combine face-to-face instruction with computer-mediated instruction (Bonk & Graham, 2006, p. 5, p. 5).

The effect of tech on the following variables

Table 6
Variables/constructs used for the evaluation of learning technologies in educational context.

Themes/aspects (No. of papers, %)[a]	Sub-theme constructs	No. of papers	%[b]	Established instruments[c]	Self-developed instruments[d]
Learning (287, 78.6%)	• Knowledge, achievement or performance	264	72.3%	32.2%	67.8%
	• Cognitive load/effort (e.g., mental effort)	42	11.5%	74.0%	26.0%
	• Skills development (e.g., interpersonal skills, motor skills, verbal and non-verbal skills or communication skills)	27	7.4%	70.6%	29.4%
	• Learning styles or learning strategies	8	2.2%	88.8%	11.2%
Affective Elements (225, 61.6%)	• Perceptions, intentions or preferences	121	33.2%	52.9%	47.1%
	• Engagement, motivation or enjoyment	112	30.7%	69.1%	30.9%
	• Attitudes, values or beliefs	63	17.3%	63.5%	36.5%
	• Emotional problems, anxiety or boredom	19	5.2%	85.0%	15.0%
	• Self-efficacy	15	4.1%	81.3%	18.8%
Behavior (125, 34.2%)	• Usage or participation	81	22.2%	16.5%	84.5%
	• Interaction, collaboration or cooperation	44	14.1%	54.5%	45.5%
	• Self-reflection, self-evaluation or self-regulation	26	7.1%	73.1%	26.9%
Design (58, 15.9%)	• Course quality, course content, course structure, resources or overall design	58	15.9%	33.9%	66.1%
Technology (43, 11.8%)	• Functionality	19	5.2%	70.8%	29.2%
	• Perceived usefulness	19	5.2%	70.0%	30.0%
	• Perceived ease of use	14	3.8%	85.7%	14.3%
	• Adoption	8	2.2%	50.0%	50.0%
	• Accessibility	2	0.5%	50.0%	50.0%
Teaching/Pedagogy (40, 11.0%)	• Pedagogical practice, teaching strategies or teaching quality/credibility	29	7.9%	54.5%	45.5%
	• Feedback	11	3.0%	36.4%	63.6%
Presence (20, 5.5%)	• Social presence, co-presence or community	17	4.7%	84.2%	15.8%
	• Presence in the environment	5	1.4%	60.0%	40.0%
Institutional Environment (5, 1.4%)	• Institutional - institutional capacity, institutional intervention, institutional policy or institutional support	4	1.1%	75.0%	25.0%
	• External environment/factors	1	0.3%	100.0%	0.0%

Presence: willingness to engage

[a] The percentages were calculated by dividing the total number of papers containing at least one of the theme's variables/constructs by the total number of reviewed papers (i.e. 365) in this study.
[b] The percentages were calculated by dividing the number of papers that investigated the variables/constructs by the total number of reviewed papers (i.e. 365) in this study.
[c] Established instruments refers to the studies that utilized established quantitative data collection instruments to evaluate the use of learning technology.
[d] Self-developed instruments means the researchers developed their own measures.

3.2. RQ1: themes/aspects examined during the evaluation of learning technology usage

The thematic analysis of variables and constructs examined during learning technology evaluation resulted in the identification of eight evaluation themes: learning, affective elements, behavior, technology, design, teaching/pedagogy, presence, and institutional environment. These themes and the 24 corresponding sub-categories or 'constructs' are shown in Table 6. A large majority of studies evaluated more than one theme/aspect. In order to avoid double counting, the overall percentage of papers that investigated each of the eight different themes has been calculated as the number of papers that include *any* of the corresponding sub-categories. Accordingly, the number and percentage of papers in each theme is less than the sum of the sub-categories because most papers evaluated more than one sub-category. For example, there was a sum of 22 papers in the sub-categories of presence (i.e. social presence, co-presence or community and presence in the environment) but the total number of papers for the main construct presence was 20. If a paper was observed to evaluate more than one theme it has been included in the calculation for each theme, in order to illustrate in how many of the 365 papers that theme appeared.

Most of the 365 papers evaluated learning in their studies (78.6%). This shows that the main concern of learning technology studies is the learning outcomes or learning achieved as a result of learning using technology. The majority of papers (72.3%) measured knowledge, achievement or performance as learning gain, and they frequently investigated the improvement of knowledge/achievement/performance before and after the learning processes, usually according to the difference between pre-test and post-test scores. Sometimes the cognitive load or effort expended during learning were evaluated (11.5%). In addition, skills development, including interpersonal, motor, verbal and non-verbal, or communication skills were measured in some studies (7.4%). Finally, learning styles or learning strategies (2.2%) were also examined in some studies.

In addition to learning, many studies aim to evaluate the affective impact of using technology in education (61.6%). Affective elements such as perceptions, intentions or preferences were frequently measured (33.2%), followed in prevalence by engagement, motivation or enjoyment (30.7%). Other affective aspects, namely attitudes, values or beliefs (17.3%) emotional problems, anxiety or boredom (5.2%) and self-efficacy (4.1%) were also evaluated in some of the papers.

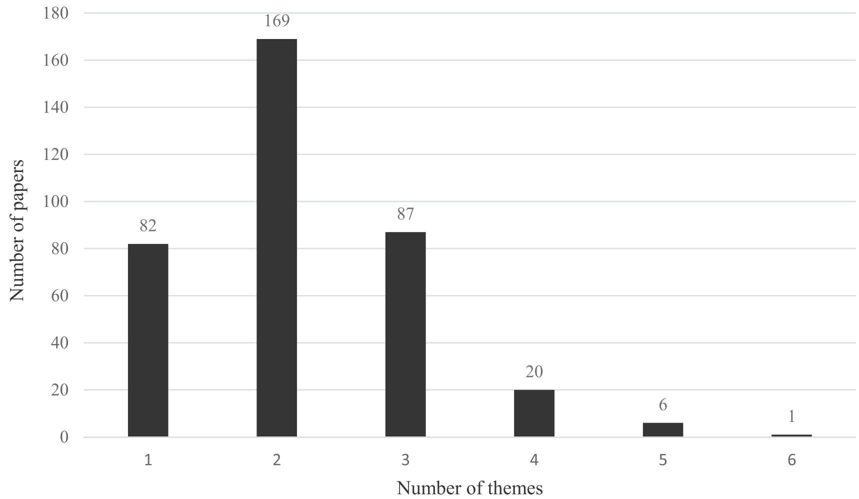

Fig. 2. Number of aspects in the reviewed papers.

Behavioral aspects of learners were often examined (34.2%). Evaluation of behavior could focus on usage or participation (22.2%), interaction, collaboration or cooperation (14.1%) as well as self-reflection, self-evaluation or self-regulation (7.1%). The design of the course was also directly evaluated in several instances, including the course quality, content, structure or overall design (15.9%).

Numerous studies examined aspects of the technology itself (11.8%). Themes investigated included functionality (5.2%), perceived usefulness (5.2%), perceived ease of use (3.8%), adoption (2.2%) and accessibility (0.5%).

Teaching or pedagogies was often evaluated as part of learning technology studies. These could be classified according to two subcategories, i.e. pedagogical practice, teaching strategies or teaching quality (7.9%) and feedback (3%). Presence was also measured in the past studies (5.5%), including social presence, co-presence or community (4.7%) or presence in the environment (1.4%). Finally, a few papers evaluated aspects of the institutional environment, such as the institutional capacity, intervention, policy or institutional support in the learning and teaching setting (1.4%).

There was a reasonable degree of variation amongst papers regarding the number of educational themes they evaluated. Fig. 2 shows the number of themes or aspects being evaluated in each study. Across the papers, 92.6% of studies performed evaluation of either one, two or three aspects. Relatively few papers (7.1%) examined four or five aspects of learning. Only the paper by Blackwell, Lauricella, and Wartella (2016) investigated six themes, namely learning, affective elements, behavior, design, technology and institutional environment. No papers performed evaluation of constructs in all eight themes.

3.3. RQ2: research methods used in the reviewed papers to evaluate the use of technology in education

The research methods used were categorized in a similar way to the systematic review done by de Klerk et al. (2015). From the 365 selected studies, quasi-experiments and experiments were by far the most commonly espoused research methods (40.5% and 23%) followed by surveys (15.6%), case studies (10.7%), mixed methods (5.5%), interviews or focus groups (1.9%), design-based research (1.6%) and observations (1.1%). See Table 7 for the range of ways that researchers described their research methods. The popularity of experiments and quasi-experiments accords with findings from the literature review by Kukulska-Hulme and Viberg

Table 7
Research methods described by authors of the reviewed papers.

Research methods	No. of papers	%
Quasi-experiments What is	148	40.5%
Experiments RCT: Randomised...	84	23.0%
Surveys	57	15.6%
Case studies	39	10.7%
Mixed methods (e.g., survey + interviews, survey + observations or survey + interviews + observations)	20	5.5%
Interviews or focus groups	7	1.9%
Design-based research	6	1.6%
Observations	4	1.1%
Total	**365**	**100.0%**

Notes: Experiments: With random assignment of participants to the experimental and control group (Cheung & Hew, 2009; Fraenkel, 2006); quasi-experiments - without random assignment of participants to experimental and control group (Cheung & Slavin, 2012).

Table 8
Types of data being collected.

Types of data	No. of papers	%
Quantitative (including Likert scale)	190	52.1%
Mixed method (both quantitative and qualitative)	148	40.5%
Qualitative	27	7.4%
Total	**365**	**100.0%**

(2018) and Connolly et al. (2012). The proportion of experiments and quasi-experiments in this *Computers and Education* sample was even higher than these other two studies, which could be seen as indicative of their analytic robustness (Boyle et al., 2016).

In terms of the actual data used for the evaluation of learning technology usage, over half of the data sources in the selected studies were quantitative (including Likert scale) in nature (Table 8). Another 40.5% of the studies collected both quantitative and qualitative data. The remaining 7.4% of papers only included qualitative data.

The large difference between the number of mixed methods studies shown in Tables 7 and 8 is explained by the fact that many of the studies that involved the collection of both quantitative and qualitative data (such as experiments, quasi-experiments, case studies, and so on) were not actually *described* as being mixed-methods investigations by their authors. The relatively small proportion of studies that are described as mixed methods by their authors is similar to results reported in the review by Zheng, Huang, and Yu (2014).

3.4. RQ3: prevailing instruments to evaluate the use of technology in education

Among the 365 papers reviewed, 243 (66.6%) applied at least one established instrument that had been used in a previous study; the remaining 122 did not use any established data collection instruments from a previous study. For instance, in their studies evaluated perceived usefulness of MOOCs, Alraimi, Zo, and Ciganek (2015) adapting the Technology Acceptance Model (Davis, 1989), and this adaptation was regarded as having used an established instrument. Alternatively, Wei, Peng, and Chou (2015) developed their own instrument to evaluate the perceived usefulness of online learning, with questions such as "*To what extent do you agree that the assignment-completion tracking function is useful for your online learning?*". Across all the constructs that were evaluated within the corpus, 62.8% were evaluated using previously established instruments (calculated using the number of papers that measured each sub-category in Table 6 and the percentage of those instances that used an established instrument, noting that only variables/constructs using quantitative data were included in the calculation). All evaluation sub-themes used established instruments in the majority of instances, except for four – the evaluation of knowledge, achievement or performance (32.2%); usage or participation (16.5%); design of the course (33.9%); and feedback (36.4%) (see final two columns of Table 6). The fact that knowledge, achievement and performance were typically evaluated using self-developed instruments was unsurprising, given the large expected variety of learning content being addressed in different studies. Usage, participation and feedback may have been evaluated using self-developed instruments because of the relatively simple or context specific nature of the evaluation being conducted (e.g., frequency counts of times an event occurred).

There were some instruments that were used repeatedly across different papers to evaluate particular aspects of technology usage in education (Table 9). For instance, to measure aspects of learning, 20 studies adapted the measurement from the Mental Effort Scale derived from Cognitive Load Theory by Paas (1992), which focuses on how much mental effort a participant invests in a particular task. Mayer's (2009) 12 principles from his Cognitive Theory of Multimedia Learning, or adaptations thereof, were used to evaluate multimedia learning in 12 papers. In addition, nine papers had adapted Bloom's (1956) Taxonomy of learning tasks (from recall through to evaluation) to design the learning activities and measure the learning performance based on the different levels of learning complexity. The Technological Pedagogical Content Knowledge (TPACK) framework (Mishra & Koehler, 2006) was adapted in nine studies, with six of them focused on evaluating the teacher's understanding of knowledge and the rest of them were evaluating the quality of teaching/pedagogical practices. The Study Processes Questionnaire (SPQ) developed by Biggs, Kember, and Leung (2001) was adapted in five papers to evaluate surface versus deep approaches to learning. The Felder-Silverman Learning Style Questionnaire developed by Felder and Silverman (Felder & Silverman, 1988) was adapted to identify individual learning styles in two papers.

In terms of evaluating technologies themselves, the Technology Acceptance Model (TAM) by Davis (1989) that measures, for instance, the perceived ease of use and perceived usefulness of technologies, was used or adapted in 12 papers. The measurement of Diffusion of Innovation Theory by Rogers (1995), which suggests five characteristics impacting on the rate of adoption of an innovation, was adapted in four studies. The Unified Theory of Acceptance and Use of Technology (UTAUT), which measures the effort expectancy, performance expectancy, social influence and facilitating conditions, by Venkatesh, Morris, Davis, and Davis (2003), was used in three papers. Bhattacherjee's Expectation–Confirmation Model (Bhattacherjee, 2001), relating to user's intention to continue using an information system, was adapted in three papers. Furthermore, two papers adapted the System Usability Scale (SUS) developed by Brooke (1996) to measure the subjective usability of a technology.

To evaluate affective elements relating to technology usage in education, there were several instruments being used and adapted across papers. For instance, researchers adapted the Motivated Strategies for Learning Questionnaire (Pintrich, Smith, Garcia, & McKeachie, 1991), in six studies, in order to evaluate the motivation and attitudes of a learning program or a class. The Instructional

Using existing instrument:
Comparability across instruments
Convenience
Reliability as validated already

Table 9
Instruments used for the evaluation of learning technologies in educational context.

Themes/aspects	Instruments*	No. of papers
Learning	• Mental Effort Scale derived from Cognitive Load Theory – Paas (1992)**	20
	• 12 principles of the Cognitive Theory of Multimedia Learning – Mayer (2009)	12
	• Taxonomy of learning tasks (from recall through to evaluation) – Bloom (1956)	9
	• Technological Pedagogical Content Knowledge (TPACK) framework – Mishra and Koehler (2006)	9
	• Study Processes Questionnaire (SPQ) – Biggs et al. (2001)	5
	• Felder-Silverman Learning Style Questionnaire – Felder and Silverman (1988)	2
Technology	• Technology Acceptance Model (TAM) – Davis (1989) Very famous, has updates	12
	• Diffusion of Innovation Theory – Rogers (1995)	4
	• Unified Theory of Acceptance and Use of Technology (UTAUT) – Venkatesh et al. (2003)	3
	• Expectation–Confirmation Model – Bhattacherjee (2001)	3
	• System Usability Scale (SUS) – Brooke (1996)	2
Affective Elements	• Motivated Strategies for Learning Questionnaire – Pintrich et al. (1991)	6
	• Instructional Materials Motivation Survey – Keller (1993)	4
	• Achievement Goal Questionnaire – Elliot and McGregor (2001)	3
	• Positive and Negative Affect Schedule (PANAS) – Watson et al. (1988)	2
	• Achievement Emotions Questionnaire (AEQ) – Pekrun et al. (2005)	2
	• Self-efficacy Scale – Wang and Hwang (2012)**	2
	• State-Trait Anxiety Inventory (STAI) – Spielberger et al. (1993)	2
	• Foreign Language Classroom Anxiety Scale (FLCAS) – Horwitz et al. (1986)	2
Presence	• Community of Inquiry Framework – Garrison et al. (2000)	4
	• Presence Questionnaire (PQ) in Virtual Environments – Witmer and Singer (1998)	2
Behavior	• Online Self-Regulated Learning Questionnaire (OSLQ) – Barnard et al. (2009)	2

Notes: *Instruments that were used in at least two of the papers.**Except for two studies (i.e. Paas, 1992; Wang & Hwang, 2012) all the previous instruments were designed by other authors.

Materials Motivation Survey (Keller, 1993) was adapted to assess the effects of the teaching approach on students' learning motivation in four papers. The Achievement Goal Questionnaire (Elliot & McGregor, 2001), which measures the achievement goals and the achievement motivation of students, was adapted in three papers. The Positive and Negative Affect Schedule (PANAS) by Watson, Clark, and Tellegen (1988) was adapted in two studies to measure the users' generalized affective state. The Achievement Emotions Questionnaire (AEQ) by Pekrun, Goetz, and Perry (2005) was adapted to measure the emotions experienced with the learning activities in two papers. The self-efficacy scale, which based on the individual self-efficacy scale developed by Wang and Hwang (2012) was adapted in two studies. Furthermore, the State-Trait Anxiety Inventory (STAI) proposed by Spielberger, Gorsuch, Lushene, Vagg, and Jacobs (1993) was adapted by two researchers to assess the learners' emotional and anxiety state. Similarly, the Foreign Language Classroom Anxiety Scale (FLCAS), measuring the learning anxiety of students in foreign language learning environment, was adapted in two studies developed by Horwitz, Horwitz, and Cope (1986).

In order to investigate the sense of presence experienced by respondents in different technology-enhanced learning environments, the Community of Inquiry Framework (Garrison, Anderson, & Archer, 2000) was adopted in four studies of asynchronous blended and online learning. The Presence Questionnaire (PQ) in Virtual Environments, measuring users' perceptions of presence in virtual environments, created by Witmer and Singer (1998), was adapted in two studies.

To evaluate the behavioral aspect of learning in technology-enhanced contexts, the Online Self-Regulated Learning Questionnaire (OSLQ), by Barnard, Lan, To, Paton, and Lai (2009), was used to evaluate the self-regulatory behavior of students in both the online and blended learning environments in two studies.

There were no instruments that were repeatedly used across studies to evaluate the design of the course or the institutional environment (as reflected in Table 9).

3.5. RQ4: relationship between the educational aspects being evaluated and the technologies, disciplines and levels of education being investigated

Examining the relationship among the eight different educational aspects being evaluated and their different contexts (i.e. technological aspects, discipline areas being taught and levels of education) provides insights into the different evaluation trends that existed within the dataset.

3.5.1. Relationship between technologies being used and the educational aspects being evaluated

Games and mobile games studies were significantly more likely to evaluate learning than studies of other technologies (63 papers, $Z = 3.403$, $p = 0.001$). For instance, Hwang, Chiu, and Chen (2015) chose to evaluate students' inquiry-based learning performance as a result of their technology-driven game-based learning approach in social studies. However, there were fewer than expected studies that focused specifically on evaluating technology aspects for game-based learning studies ($Z = -2.052$, $p = 0.040$). Only three studies, i.e. Joo-Nagata, Martinez Abad, García-Bermejo Giner, and García-Peñalvo (2017), Martín-SanJosé, Juan, Seguí, and García-García (2015) and Darban, Kwak, Deng, Srite, and Lee (2016) undertook evaluation of the technologies themselves, which was surprising given that many games and mobile games studies developed technologies for their studies.

There were proportionally less Web 2.0 studies that evaluated learning (32 papers, $Z = -3.762$, $p = 0.000$), which was also surprising, given their increasingly popular use to facilitate learning in educational contexts. As a counter-example Orús et al. (2016) studied the effects of learner-generated videos in YouTube on learning outcomes.

Studies involving online discussion platforms more frequently evaluated behaviors, especially the interactivity between learners (11 papers, $Z = 2.463$, $p = 0.014$). For instance, Yücel and Usluel (2016) evaluated the interaction of learners in an online collaborative learning environment. Evaluation of behaviors in forums may be due to the proposition that well-designed online discussion forums should be able to facilitate better participation and interaction (Zheng & Warschauer, 2015).

Studies of mobile learning, including utilizing tablets, iPads, computers, interactive tools or mobile devices in learning and teaching, were more likely to evaluate teaching strategies and pedagogy than studies of other technologies (11 papers, $Z = 2.399$, $p = 0.016$). For instance, Engin and Donanci (2015) studied the pedagogy surrounding the impact of iPad use in English for Academic Purposes (EAP) classes at university. Such studies help to investigate how teachers are now adjusting and adapting their teaching strategies and pedagogy to build their own capacities in accordance with technology integration (Tondeur, Pareja Roblin, van Braak, Voogt, & Prestridge, 2017).

Online learning course delivery and e-learning studies were more likely to evaluate presence ($Z = 2.385$, $p = 0.017$). For instance, Luo, Zhang, and Qi (2017) investigated the learners' sense of membership in e-learning environment. Presence has been shown to impact student motivation and participation, actual and perceived learning, course and instructor satisfaction, and retention in online courses (Richardson, Maeda, Lv, & Caskurlu, 2017), so the emphasis upon investigating presence in online courses is understandable.

3.5.2. Relationship between disciplines and educational aspects being evaluated

Science-related subjects were more likely to measure learning performance (91 papers, $Z = 3.072$, $p = 0.002$). As compared to non-science subjects, it may be easier to measure learning outcomes objectively in science (Strimaitis, Schellinger, Jones, Grooms, & Sampson, 2014). However, relatively few science-related subjects evaluated affective elements (54 papers, $Z = -2.129$, $p = 0.033$). This may also be due to the generally more objective focus of science-related disciplines.

In the business and economics disciplines, more than expected amount of papers evaluated design aspects of the technology-enhanced learning approach (7 papers in total, $Z = 2.404$, $p = 0.016$). For instance, Orús et al. (2016) evaluated the course quality in an *Introduction to Marketing* course. In general, no statistical significantly trend was identified between the evaluation of behavior and the discipline areas.

3.5.3. Relationship between levels of education and the educational aspects being evaluated

Fewer than expected studies evaluated learning with educational technology in the university/college environment (131 papers, $Z = -2.089$, $p = 0.037$). A typical example of a study that did was the study by Hanus and Fox (2015) that evaluated the effects of gamification in the classroom by measuring academic performance in a large USA university. The lower proportion of higher education studies that evaluate learning may be due to the fact that researchers in universities use convenience samples to measure other aspects of education (for instance, presence, design, affective elements).

Studies focused on a mix of educational levels were more likely to evaluate the teaching/pedagogy aspects of education (5 papers, $Z = 3.233$, $p = 0.001$). For instance, research conducted by Olofson, Swallow, and Neumann (2016) measured teaching practices in both primary and secondary schools, using multiple case studies of primary and secondary school teachers to illustrate different perspectives. Such studies focusing on multi-levels of education serve to demonstrate whether pedagogies are generalizable to multiple age groups (Johnson & Onwuegbuzie, 2004).

Researchers more frequently evaluated presence in adult learning environments (4 papers, $Z = 2.395$, $p = 0.017$). Luo et al. (2017) investigated the student's sense of community in e-learning environment. Ke (2010) examined the presence aspect of adult learners in ten WebCT-based online courses. In these studies, authors tend to argue that the sense of presence and virtual relationships developed among adult students forms a sense of community that in turn contributes to the overall success of distance education programs.

In general, the relationship between the identified evaluation themes and the different educational technology contexts, disciplines and levels of education was various. For instance, mobile games tended to place more emphasis on evaluating learning outcomes as compared to Web 2.0 technologies. Learning aspects were more likely to be evaluated in science subjects. The results from this section will be further discussed in the Conclusion, noting that further research would be required to confidently determine the reasons for the different evaluation tendencies within the different contexts.

4. Discussion

4.1. Implications and critical reflections

The fact that studies across the technology-enhanced learning research papers focused on the evaluating learning, affective elements, behaviors, design, technology, pedagogy, presence and institutional environment, was not in and of itself surprising (ref. Table 6). However, the proportion of studies focusing on each aspect was in some cases unexpected by the research team. For instance, only 1.4% of the papers evaluated 'institutional environment'. McGill, Klobas, and Renzi (2014) argue that institutional supports, e.g. financial support, technical support and training, formal recognition of using learning technology in teaching, are critical for the success of technology-enhanced education. As well, most studies focused on evaluating one or two aspects of education

(Fig. 2), with only 7% of studies analyzed focusing on four, five, or six aspects of education. This raises the important question of whether educators are evaluating the use of technology in education from an holistic perspective. It may be that in order to focus their study, researchers concentrate on specific aspects of the education that they deem relevant for their research questions. On the other hand, it is possible that focusing on evaluating only some aspects of education is a consequence of a certain degree of analytical narrowness. We hope that our study may encourage researchers to at least deliberately reflect upon the different aspects of education that they evaluate in their technology-enhanced learning studies, and in some circumstances broaden the range of aspects that they consider.

There were notable similarities between this review of evaluation practices and other educational technology reviews in terms of the subject areas, education levels, technologies and methods that were featured. For instance, science is the most often evaluated subject (e.g., Crompton et al., 2017; Hwang & Tsai, 2011; Liu et al., 2014, etc.). There is a variety of educational levels, contexts and participants associated with the technology-enhanced learning (Sung et al., 2016). The review by Boyle et al. (2016) identified the surge of interest in the use of games for educational purposes. This similarity tends to indicate commonalities between the characteristics of the sample selected for this study and the educational technology literature more broadly. This in turn, along with the large sample size, could also be seen to support the generalizability of the results, however, this is open to individual interpretation.

We note the relatively high proportion of purely experimental and quantitative studies that have been conducted in recent years. While acknowledging the important contributions that these studies can make in identifying effects, we encourage the concurrent use of qualitative methods to offer the field explanations of 'why' and 'how' certain phenomena take place. The wide variety of evaluation instruments being used across the learning technology field, including self-developed and often un-validated instruments, also highlights an important issue. While not wanting to pigeon-hole studies, or advocate the use of inappropriate evaluation instruments, use of inconsistent approaches to evaluation across studies means that it is more difficult for the field to accurately and reliably compare findings, and hence make relative judgements. More frequent use of common evaluation instrumentation could help the field to better interrelate findings and more confidently qualify effects.

The fact that trends were observed in the way in which the use of technologies were evaluated for different technologies, disciplines, levels, highlights that context drives evaluation, at least to some extent. Whether this is due to limitations in how researchers and educators within certain contexts view the use of technology, or alternately because good contextualized evaluation demands certain evaluation processes, is an issue for consideration. For instance, the fact that studies in higher education did not evaluate learning as much as studies at other levels, may be for convenience reasons. Do the sciences tend to overlook the study of affective elements in favor of measuring learning outcomes because the latter is more akin to the objective approaches of their discipline? These questions are posed for critical reflection and further research rather than as definitive conclusions.

4.2. Applications and utility

The findings from this study can be used in several ways. Firstly, the findings characterized the way in which usage of technology in education has been evaluated throughout the field, with the percentages of each theme and sub-category indicating the relative emphasis being placed on the various possible constructs. This constitutes new knowledge for the field – while previous review studies had identified some areas of educational technology evaluation, none of the previous studies provided such a detailed and systematic analysis of the sorts of evaluation processes occurring when technology is being used for learning purposes, classified by specific sub-categories of evaluation. As well, this is the first study that broadly details the evaluation of technology usage in education across disciplines, levels of learning, and technologies.

Secondly, the areas that emerged from the analysis of learning technology evaluation literature provide the field with a framework of constructs that can be considered when researchers and educators are conducting evaluations of learning technology usage. For instance, a researcher can reflect upon whether learning, affective elements, behaviors, design, technology, pedagogy, presence, or institutional environment are important to their study, and if they are, consider which sub-categories may warrant investigation. This also offers educators a range of aspects that they may choose to focus upon in the creation of their learning designs, and some educators may choose to reflect upon whether their designs address all eight aspects of education that emerged from the study.

Thirdly, the identification of prevailing instruments being used to evaluate the various aspects of technology usage in educational contexts, provides educators and researchers with a valuable starting point for investigating and selecting the instruments that they may use in their studies. When embarking on learning technology evaluation it is often useful to know which instruments are most typically used to evaluate different aspects of educational technology usage, and to that extent this paper provides guidance for various aspects of educational technology evaluation in one source.

Fourthly, the framework of themes and sub-categories could potentially form the basis of new evaluation instruments, or even a generalized model for evaluating technology usage in education. For instance, the themes could constitute initial factors, and the sub-categories can form the components of those factors. This is an area of future work that is being considered by the research team.

Finally, educational institutions may choose to use the findings of this study to design the evaluation of their learning technology usage, including the different aspects of education that they may consider, emphasize, and compare, and the instrumentation they use, depending on their institutional priorities. Similar approaches could also be applied by governments, who may also choose to use the results from this study to encourage standardized and holistic evaluation of technology usage in their educational programs and grants.

4.3. Limitations – threats to validity

The most notable limitation of this study was that it only drew from a single journal. Drawing papers from the highest quality journal in the field was a logical decision considering the large number of papers that needed to be analyzed according to the broad inclusion criteria. This way we could focus our study on high quality evaluation of learning technology use. Nevertheless, we cannot discount the possibility that quite different results would have been achieved if a different journal was selected. To mitigate this, we do note that the results of this study broadly concur with findings from other reviews, which indicates that the results may be somewhat representative of the field more generally. In any case, the main point of this study was to establish the sorts of evaluations taking place in educational technology research, rather than to determine exact proportions across the entire field. To this extent we contend that the sampling and analysis process enabled us to address the research questions. We encourage readers to consider how they apply the findings from this study and the extent to which the findings are transferable to their context.

Another limitation is that the study only drew from a three-year time frame, from 2015 to 2017, meaning that the findings cannot be claimed to represent evaluation within the field over an extended history. Limiting the time frame being investigated was once again a consequence of needing to constrain the scope of the analysis, which in any case resulted in a large sample of 365 papers. A positive consequence of constraining the investigation to recent years was that the analysis is more relevant to current evaluation practices.

5. Conclusion

This paper presented findings from systematic review of the evaluation of technology usage in education based on 365 papers published in Computers & Education between 2015 and 2017. The papers analyzed encompassed a diverse range of education levels, disciplines, technologies (as outlined in Table 3, Table 4, and Table 5, respectively). Through thematic analysis of the literature this study found that empirically, the evaluation of learning technology use tends to focus on learning (78.6%), affective elements (61.6%), behaviors (34.2%), design (15.9%), technology (11.8%), pedagogy (11%), presence (5.5%) and institutional environment (1.4%). The analysis identified several sub-themes that constituted these eight overarching themes (outlined in Table 6). Most studies adopted quasi-experimental methodologies and experiments (40.5% and 23% respectively), however there was a substantial proportion of studies that involved both qualitative and quantitative data sources (40.5%), even if they were not named as mixed-methods studies (see Tables 7 and 8). The majority of studies used established instruments for collection of quantitative data (66.6%), implying that a large minority did not, and there was a wide variety of instruments being used across different studies to evaluate different aspects of learning during technology usage.

The research was also able to characterize different evaluation emphases in different educational technology contexts. Studies of mobile games tended to more frequently evaluate learning but less frequently evaluate technology aspects. Web 2.0 technologies tended to place less emphasis on evaluating learning outcomes than other technologies, whereas discussion forums were more likely to focus on evaluating interaction. Mobile learning studies tended to more often evaluate teaching and pedagogy, and studies of online-learning environments more commonly evaluated presence. In terms of disciplines, studies in the Science context more frequently evaluated learning outcomes, but less frequently evaluated affective aspects of learning. Business and Economics studies were more likely to evaluate overall design aspects of their interventions. At the adult education level there was a greater tendency to evaluate presence than at other levels, university and college studies were less likely to examine learning, and teaching and pedagogy was more frequently examined in studies that focused upon mixed levels of education.

There are evidently a large variety of contexts, methodologies and perspectives that have been used to evaluate the use of technology in education, including many discipline, level, and technology specific trends. The large amount of literature across many different journals in the educational technology field has historically made it challenging to embark on educational technology research and evaluation from an informed perspective. By analyzing the aspects of education, research methods and instruments used in a large sample of high quality learning technology studies, and by identifying the contextual trends that arose for different educational levels, discipline areas and technologies, we intend that this paper supports researchers, teachers, and the broader education sector to more effectively and efficiently design and evaluate use of technology in education.

Declarations of interest

None.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2019.01.010.

References

Alraimi, K. M., Zo, H., & Ciganek, A. P. (2015). Understanding the MOOCs continuance: The role of openness and reputation. *Computers & Education, 80*, 28–38.

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk (2014). Augmented reality trends in education: A systematic review of research and applications. *Journal of Educational Technology & Society, 17*(4), 133–149.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

Baran, E. (2014). A review of research on mobile learning in teacher education. *Journal of Educational Technology & Society, 17*(4), 17–32.

Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S.-L. (2009). Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education, 12*(1), 1–6.

Beckers, J., Dolmans, D., & van Merriënboer, J. (2016). e-Portfolios enhancing students' self-directed learning: A systematic review of influencing factors. *Australasian Journal of Educational Technology, 32*(2), 32–46.

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.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly, 25*(3), 351–370.

Biggs, J., Kember, D., & Leung, D. Y. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133–149.

Blackwell, C. K., Lauricella, A. R., & Wartella, E. (2016). The influence of TPACK contextual factors on early childhood educators' tablet computer use. *Computers & Education, 98*, 57–69.

Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook 1: The cognitive domain.* New York: David McKay Co Inc.

Bonk, C. J., & Graham, C. R. (2006). *The handbook of blended learning: Global perspectives, local designs* (1st ed.). San Francisco, CA: Pfeiffer Publishing.

Bower, M. (2016). Deriving a typology of Web 2.0 learning technologies. *British Journal of Educational Technology, 47*(4), 763–777.

Bower, M. (2017). Technology integration as an educational imperative. In M. Bower (Ed.). *Design of technology-enhanced learning: Integrating research and practice* (pp. 1–16). Bingley: Emerald Publishing Limited.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Thousand Oaks: Sage Publications.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., ... Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education, 94*, 178–192.

Brooke, J. (1996). SUS-A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.). *Usability ealuation in industry* (pp. 189–194). London: Taylor & Francis.

Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education, 58*, 154–161.

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. *Computers & Education, 87*, 396–422.

Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education, 105*, 14–30.

Cheng, M.-T., Chen, J.-H., Chu, S.-J., & Chen, S.-Y. (2015a). The use of serious games in science education: A review of selected empirical research from 2002 to 2013. *Journal of Computers in Education, 2*(3), 353–375.

Cheng, M.-T., Lin, Y.-W., & She, H.-C. (2015b). Learning through playing Virtual Age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education, 86*, 18–29.

Cheung, W. S., & Hew, K. F. (2009). A review of research methodologies used in studies on mobile handheld devices in K-12 and higher education settings. *Australasian Journal of Educational Technology, 25*(2), 153–183.

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.

Chun, D., Kern, R., & Smith, B. (2016). Technology in language use, language teaching, and language learning. *The Modern Language Journal, 100*(S1), 64–80.

Clarivate Analytics (2019). *2018 Journal citation reports*Philadelphia, PA: Clarivate Analytics.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122.

Claros, I., Cobos, R., & Collazos, C. A. (2016). An approach based on social network analysis applied to a collaborative learning experience. *IEEE Transactions on Learning Technologies, 9*(2), 190–195.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686.

Conole, G., De Laat, M., Dillon, T., & Darby, J. (2008). 'Disruptive technologies', 'pedagogical innovation': What's new? Findings from an in-depth study of students' use and perception of technology. *Computers & Education, 50*, 511–524.

Cook, D. A., Brydges, R., Hamstra, S. J., Zendejas, B., Szostek, J. H., Wang, A. T., ... Hatala, R. (2012). Comparative effectiveness of technology-enhanced simulation versus other instructional methods: A systematic review and meta-analysis. *Simulation in Healthcare, 7*(5), 308–320.

Crompton, H., Burke, D., & Gregory, K. H. (2017). The use of mobile learning in PK-12 education: A systematic review. *Computers & Education, 110*, 51–63.

Darban, M., Kwak, D.-H., Deng, S., Srite, M., & Lee, S. (2016). Antecedents and consequences of perceived knowledge update in the context of an ERP simulation game: A multi-level perspective. *Computers & Education, 103*, 87–98.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Ekanayake, S. Y., & Wishart, J. (2015). Integrating mobile phones into teaching and learning: A case study of teacher training through professional development workshops. *British Journal of Educational Technology, 46*(1), 173–189.

Elliot, A. J., & McGregor, H. A. (2001). A 2× 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501.

Engin, M., & Donanci, S. (2015). Dialogic teaching and iPads in the EAP classroom. *Computers & Education, 88*, 268–279.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education, 78*(7), 674–681.

Foshee, C. M., Elliott, S. N., & Atkinson, R. K. (2016). Technology-enhanced learning in college mathematics remediation. *British Journal of Educational Technology, 47*(5), 893–905.

Fraenkel, J. R. (2006). *How to design and evaluate research in education* (6th ed.). Boston: McGraw-Hill.

Gao, F., Luo, T., & Zhang, K. (2012). Tweeting for learning: A critical analysis of research on microblogging in education published in 2008–2011. *British Journal of Educational Technology, 43*(5), 783–801.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105.

Google (2019). *Google scholar metrics.* Retrieved January 16, 2019https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=soc_educationaltechnology.

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education, 102*, 202–223.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80*, 152–161.

Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. *Journal of Research on Technology in Education, 41*(4), 393–416.

Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology, 41*(1), 33–55.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132.

Hwang, G.-J., Chiu, L.-Y., & Chen, C.-H. (2015). A contextual game-based learning approach to improving students' inquiry-based learning performance in social

studies courses. *Computers & Education, 81*, 13–25.
Hwang, G.-J., & Tsai, C.-C. (2011). Research trends in mobile and ubiquitous learning: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology, 42*(4), E65–E70.
Iriti, J., Bickel, W., Schunn, C., & Stein, M. K. (2016). Maximizing research and development resources: Identifying and testing "load-bearing conditions" for educational technology innovations. *Educational Technology Research & Development, 64*(2), 245–262.
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14–26.
Joo-Nagata, J., Martinez Abad, F., García-Bermejo Giner, J., & García-Peñalvo, F. J. (2017). Augmented reality and pedestrian navigation through its implementation in m-learning and e-learning: Evaluation of an educational program in Chile. *Computers & Education, 111*, 1–17.
Ke, F. (2010). Examining online teaching, cognitive, and social presence for adult students. *Computers & Education, 55*(2), 808–820.
Keller, J. M. (1993). *Manual for instructional materials motivational survey (IMMS).* Tallahassee, Florida: Florida State University.
Kennedy, J. (2014). Characteristics of massive open online courses (MOOCs): A research review, 2009-2012. *The Journal of Interactive Online Learning, 13*(1), 1–16.
Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: What is 'enhanced'and how do we know? A critical literature review. *Learning, Media and Technology, 39*(1), 6–36.
Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering version 2.3, technical report EBSE-2007-01*Keele and Durham: Keele University and University of Durham.
Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O. P., Turner, M., Niazi, M., et al. (2010). Systematic literature reviews in software engineering – a tertiary study. *Information and Software Technology, 52*(8), 792–805.
de Klerk, S., Veldkamp, B. P., & Eggen, T. J. H. M. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education, 85*, 23–34.
Kordaki, M., & Gousiou, A. (2017). Digital card games in education: A ten year systematic review. *Computers & Education, 109*, 122–161.
Kukulska-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology, 49*(2), 207–218.
Lee, H. S., Linn, M. C., Varma, K., & Liu, O. L. (2010). How do technology-enhanced inquiry science units impact classroom learning? *Journal of Research in Science Teaching, 47*(1), 71–90.
Liu, M., Scordino, R., Geurtz, R., Navarrete, C., Ko, Y., & Lim, M. (2014). A look at research on mobile learning in K–12 education from 2007 to the present. *Journal of Research on Technology in Education, 46*(4), 325–372.
Looi, C. K., Zhang, B., Chen, W., Seow, P., Chia, G., Norris, C., et al. (2011). 1: 1 mobile inquiry learning experience for primary science students: A study of learning effectiveness. *Journal of Computer Assisted Learning, 27*(3), 269–287.
Luo, N., Zhang, M., & Qi, D. (2017). Effects of different interactions on students' sense of community in e-learning environment. *Computers & Education, 115*, 153–160.
Manca, S., & Ranieri, M. (2013). Is it a tool suitable for learning? A critical review of the literature on Facebook as a technology-enhanced learning environment. *Journal of Computer Assisted Learning, 29*(6), 487–504.
Martín-SanJosé, J.-F., Juan, M. C., Seguí, I., & García-García, I. (2015). The effects of computer-based games and collaboration in large groups vs. collaboration in pairs or traditional methods. *Computers & Education, 87*, 42–54.
Mayer, R. E. (2009). *Multimedia learning* (2 ed.). New York, NY, USA: Cambridge University Press.
McGill, T. J., Klobas, J. E., & Renzi, S. (2014). Critical success factors for the continuation of e-learning initiatives. *The Internet and Higher Education, 22*, 24–36.
Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* Retrieved from Washington, D.C www.ed.gov/about/offices/list/opepd/ppss/reports.html.
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 & Education, 70*, 29–40.
Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education, 56*, 769–780.
Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *British Medical Journal, 339*(July), 1–8.
Mohsen, M. A. (2016). The use of help options in multimedia listening environments to aid language learning: A review. *British Journal of Educational Technology, 47*(6), 1232–1242.
Oliver, M. (2000). An introduction to the evaluation of learning technology. *Educational Technology & Society, 3*(4), 20–30.
Olofson, M. W., Swallow, M. J. C., & Neumann, M. D. (2016). TPACKing: A constructivist framing of TPACK to analyze teachers' construction of knowledge. *Computers & Education, 95*, 188–201.
Orús, C., Barlés, M. J., Belanche, D., Casaló, L., Fraj, E., & Gurrea, R. (2016). The effects of learner-generated videos for YouTube on learning outcomes and satisfaction. *Computers & Education, 95*, 254–269.
Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434.
Pegrum, M., Howitt, C., & Striepe, M. (2013). Learning to take the tablet: How pre-service teachers use iPads to facilitate their learning. *Australasian Journal of Educational Technology, 29*(4), 464–479.
Pekrun, R., Goetz, T., & Perry, R. P. (2005). *Academic emotions questionnaire-mathematics (AEQ-M): User's manual.* Munich: University of Munich.
Pérez-Sanagustín, M., Nussbaum, M., Hilliger, I., Alario-Hoyos, C., Heller, R. S., Twining, P., & Tsai, C.-C. (2017). Research on ICT in K-12 schools – a review of experimental and survey-based studies in computers & education 2011 to 2015. *Computers & Education, 104*, A1–A15.
Petri, G., & Gresse von Wangenheim, C. (2017). How games for computing education are evaluated? A systematic literature review. *Computers & Education, 107*, 68–90.
Phillips, R., McNaught, C., & Kennedy, G. (2012). *Evaluating e-learning: Guiding research and practice.* New York: Routledge.
Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning Questionnaire (MSLQ).* Michigan: The University of Michigan.
Prior, D. D., Mazanov, J., Meacheam, D., Heaslip, G., & Hanson, J. (2016). Attitude, digital literacy and self efficacy: Flow-on effects for online learning behavior. *The Internet and Higher Education, 29*, 91–97.
Rea, L. M., & Parker, R. A. (2014). *Designing and conducting survey research: A comprehensive guide.* San Francisco: John Wiley & Sons.
Richardson, J. C., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior, 71*, 402–417.
Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: The Free Press.
Salleh, N., Mendes, E., & Grundy, J. (2011). Empirical studies of pair programming for CS/SE teaching in higher education: A systematic literature review. *IEEE Transactions on Software Engineering, 37*(4), 509–525.
Smetana, L. K., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education, 34*(9), 1337–1370.
Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1993). *Manual for the state-trait anxiety inventory.* Palo Alto, California: Consulting Psychologists Press.
Strimaitis, A. M., Schellinger, J., Jones, A., Grooms, J., & Sampson, V. (2014). Development of an instrument to assess student knowledge necessary to critically evaluate scientific claims in the popular media. *Journal of College Science Teaching, 43*(5), 55–68.
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.
Tess, P. A. (2013). The role of social media in higher education classes (real and virtual) – a literature review. *Computers in Human Behavior, 29*, A60–A68.
Tondeur, J., Pareja Roblin, N., van Braak, J., Voogt, J., & Prestridge, S. (2017). Preparing beginning teachers for technology integration in education: Ready for take-off? *Technology, Pedagogy and Education, 26*(2), 157–177.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.
Wang, S. L., & Hwang, G. J. (2012). The role of collective efficacy, cognitive quality, and task cohesion in computer-supported collaborative learning (CSCL). *Computers & Education, 58*(2), 679–687.
Wang, C.-H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education, 34*(3), 302–323.
Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063.
Wei, H.-C., Peng, H., & Chou, C. (2015). Can more interactivity improve learning achievement in an online course? Effects of college students' perception and actual use of a course-management system on their learning achievement. *Computers & Education, 83*, 10–21.
Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.
Wollscheid, S., Sjaastad, J., & Tømte, C. (2016). The impact of digital devices vs. Pen(cil) and paper on primary school students' writing skills – a research review. *Computers & Education, 95*, 19–35.
Wong, L.-H., & Looi, C.-K. (2011). What seams do we remove in mobile-assisted seamless learning? A critical review of the literature. *Computers & Education, 57*, 2364–2381.
Wu, W.-H., Jim Wu, Y.-C., Chen, C.-Y., Kao, H.-Y., Lin, C.-H., & Huang, S.-H. (2012). Review of trends from mobile learning studies: A meta-analysis. *Computers & Education, 59*, 817–827.
Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., ... Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89.
Yücel, Ü. A., & Usluel, Y. K. (2016). Knowledge building and the quantity, content and quality of the interaction and participation of students in an online collaborative learning environment. *Computers & Education, 97*, 31–48.
Zheng, L., Huang, R., & Yu, J. (2014). Identifying computer-supported collaborative learning (CSCL) research in selected journals published from 2003 to 2012: A content analysis of research topics and issues. *Educational Technology & Society, 17*(4), 335–351.
Zheng, B., & Warschauer, M. (2015). Participation, interaction, and academic achievement in an online discussion environment. *Computers & Education, 84*, 78–89.