Received: 23 October 2020 | Revised: 10 August 2021 | Accepted: 22 December 2021

DOI: 10.1111/jcal.12645

ARTICLE

Journal of Computer Assisted Learning WILEY

What should we evaluate when we use technology in education?

Follow-up study on their previous work Lai.Bower-19, confirm it is the validity.

Jennifer W. M. Lai[1] | Matt Bower[1] | John De Nobile[1] | Yvonne Breyer[2]

[1]Macquarie School of Education, Macquarie University, Sydney, New South Wales, Australia

[2]Macquarie Business School, Macquarie University, Sydney, New South Wales, Australia

Correspondence
Jennifer W. M. Lai, Macquarie School of Education, Macquarie University, Sydney, NSW 2109, Australia.
Email: jennifer.lai@mq.edu.au

Abstract

Background: There is a lack of critical or empirical work interrogating the nature and purpose of evaluating technology use in education.

Objectives: In this study, we examine the values underpinning the evaluation of technology use in education through field specialist perceptions. The study also poses critical reflections about the rigour of evaluation instruments development in the educational technology field.

Method: A total of 48 domain specialists were surveyed to investigate the face and content validity of 39 items under eight constructs, with both qualitative and quantitative data from the survey analysed.

face: does it make sense to understand
content: does it make sense to use

Results: There was an alignment between the constructs that the specialists felt were important and the constructs that have typically been the focus of empirical studies, with field specialists indicating high relevance scores for technology ($M = 3.24/4$), learning outcomes ($M = 3.20/4$), affective elements ($M = 3.19/4$), behaviour ($M = 3.15/4$), presence/community ($M = 3.07/4$), teaching/pedagogy ($M = 3.01/4$), design ($M = 2.96/4$), and institutional environment ($M = 2.86/4$). Only a minority of other studies were found to perform face and content validity checks and even then only with small samples of respondents (usually $n \leq 5$).

Implications: Specialists in educational technology research confirm that all eight dimensions are important to consider when evaluating the use of technology in education. Thorough face and content validity processes should be adopted when developing educational technology evaluation instruments. Further work has validated an eight factor 28 item instrument for evaluating the use of technology in education using a large sample of students from a global open learning online course.

KEYWORDS

cultural and social implications, evaluation methodologies, pedagogical issues, teaching/learning strategies

1 | INTRODUCTION

Learning technology (interchangeably termed instructional or educational technology) and information technology have rapidly evolved over recent decades to offer new ways of supporting learning and teaching (Tharp & Chamberlain, 2016; Xie et al., 2019). There is an enormous variety of technologies at the educator's disposal, for instance, mobile games, Web2.0 learning technologies (e.g., social media, Social Networking Systems, wikis or blogs), virtual worlds and virtual reality, digital instructions or visual aids, management systems (e.g., classroom management systems, learning management systems or self-regulated learning systems), online discussion platforms, technology-enhanced feedback systems such as student responses systems, MOOCs, and so on. This variety of technologies coupled with the multifaceted resourcing (Carver, 2016), time (Arpaci, 2017), and institutional considerations (Porter et al., 2014) associated with their use makes effective evaluation of technologies a complex endeavour. However, comprehensive and effective evaluation of technology use is essential for educators, researchers, and institutional leaders to make accurate decisions in relation to learning technology practices and strategies (Alqahtani & Rajkhan, 2020; Yang & Huang, 2015).

There are many definitions of evaluation in the literature that make it difficult to find a universally accepted view of what it entails. Evaluation can be fundamentally defined as the process by which people make judgements about value and worth (Oliver, 2000). Evaluation has also been characterized as a procedure to provide information to make decisions about the product or process being investigated (Phillips, McNaught et al., 2012). In the context of learning technology, the evaluation and judgement process is complex, multifaceted, and often controversial (Oliver, 2000). For instance, the evaluations of the effectiveness of computer technology usage in schools and homes often yield mixed evidence, with many null effects (Bulman & Fairlie, 2016). There is consensus that developing new ways and combinations of methods to investigate and evaluate technology-assisted learning is of ultimate benefit to educators and researchers (Cox, 2013; Iriti et al., 2016).

Previously, researchers have proposed that there is insufficient understanding about various factors that relate to the evaluation of technology integration in educational settings (Inan & Lowther, 2010; Kirschner & Kester, 2016). Phillips, Kennedy et al. (2012) point out that investigations of learning technology can involve a complex mixture of variables within the evaluation process, though fall short of empirically defining what these variables are. To this extent, a recent systematic literature review examining the ways that the use of technology has been evaluated in education found that the evaluation of learning technology use can be categorized into eight themes: *learning outcomes*, *affective elements*, *behaviours*, *design*, *technology elements*, *pedagogy*, *presence/community*, and *institutional environment* (Lai & Bower, 2019). Yet, as well as what has been evaluated, there is also a subjective aspect of educational technology evaluation that relates to the question, 'what *should* be evaluated'.

The present study aims to advance the extant literature and provide a deeper understanding of the evaluation of educational technology in teaching and learning by examining which elements and dimensions accomplished researchers within international educational technology community regard as most important to consider when evaluating the use of technology in education. Specifically, it addresses the following main research question (*RQ1*): *Which elements and dimensions do accomplished researchers within international educational technology community regard as most important to consider when evaluating the use of technology in education?* Additionally, we also address the following related questions. *RQ2*: *What is the alignment between the perceptions of field specialists and previous empirical work?* *RQ3*: *What are the face and content validity of specific questions based on the feedback of the field specialists?* This study also poses critical reflections about the rigour of evaluation instrument development in the educational technology field based on an analysis of processes typically adopted.

2 | BACKGROUND

2.1 | Theoretical perspectives of educational technology evaluation

As a background to understand what we should evaluate when we use technology in education, it is worthwhile to first broadly consider how pertinent theoretical models and frameworks relate to the evaluation of learning technology use. Several general theories have been used in the educational research field to explain activities and interactions. For instance, Activity Theory provides a general framework for analysing interactions between subject and object, in which mutual transformations are accomplished. This interaction is usually mediated by physical tools (computers) or mental tools (notations, maps), which shape the way humans interact with the world (Kaptelinin & Nardi, 2007). This theory has been used, for instance, to investigate technology-mediated learning through Facebook (Rambe, 2012), evaluate the user experience of adaptive digital educational games (Law & Sun, 2012), examine an adaptive e-learning system for teaching Geography (Peña-Ayala et al., 2014), study the implementation of a technology-supported distance firefighter training programme (Holmgren et al., 2017) and analyse interactions in the flipped classroom (Zheng et al., 2020). Roger's Diffusion of Innovation theory (1995) is another theoretical referent that explains how innovation diffuses in a social system and enables researchers to evaluate, especially the adoption of new technology in the educational context. The theory has been used to understand blended learning adoption in higher education (Porter et al., 2016), examine university instructors' perceptions of technology integration (Ashrafzadeh & Sayadian, 2015), investigate the adoption of game-based learning in Taiwan (Li & Huang, 2016), or study the secondary teachers' decisions to adopt geospatial technologies in classroom (Curtis, 2020). Institutional theory points out that beliefs, attitudes, and behaviours of individuals and organizations are strongly influenced by various networks and interactions in a social context (Scott, 2014). Researchers applied this theory to investigate e-learning adoption (Jan et al., 2012), examine the effectiveness of assessment technologies (Burch, 2010) or assess the lecturers' acceptance towards online learning system in South East Asia (Kusuma et al., 2020). However, these theories in and of themselves do not relate specifically to the educational technology field and do not provide direct guidance on how to evaluate the use of technology in education, resulting in a potential disconnect between theory and educational technology evaluation practices.

There are instruments that have been developed for users' acceptance of technology, for instance, the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2). The TAM model was developed from Theory of

Reasoned Action by Davis (1989) to predict acceptance and future usage of technological innovation, with factors such as 'ease of use' and 'usefulness' used to determine the degree to which a technology is accepted by individuals. In current learning technology studies about technology adoption, TAM is the most common theory being used to explain technology acceptance in various contexts (Briz-Ponce et al., 2017; del Barrio-Garcia et al., 2015; Ranellucci et al., 2020; Shroff et al., 2011). However, some researchers criticized the TAM framework for neglecting to investigate other essential predictors and factors that may affect the adoption and acceptance of technology (Tarhini et al., 2014). Venkatesh et al. (2003) subsequently derived a unified theoretical model to explain behavioural intention to use technology—the Unified Theory of Acceptance and Usage Theory (UTAUT)—that captured the essential elements of eight previously established models with factors including performance expectancy, effort expectancy, social influence and facilitating conditions. It was argued that UTAUT may not be suitable for users whose intention to use technology is influenced by a broader range of factors (Venkatesh et al., 2012), hence another model was developed (UTAUT2) that also incorporated factors of hedonic motivation, the price value and the habit of using the technology (Venkatesh et al., 2012). Yet, research applying the UTAUT and UTAUT2 is still in its relatively early stages of development, with no clear areas of maturity (Tamilmani et al., 2017). Furthermore, these models mainly focus on technology adoption or acceptance but comparatively less on the evaluation of technology use in *education* with respect to pertinent aspects such as learning outcomes, pedagogy, learning design, and so on.

There are educational technology-specific frameworks that have been used for evaluation purposes. For instance, the Technological Pedagogical Content Knowledge (TPACK) framework by Mishra and Koehler (2006) has been used extensively as a way of evaluating teacher knowledge. For example, Baser et al. (2016) have used TPACK to assess preservice teachers learning to teach English as a foreign language (EFL). Deng et al. (2017) have applied the framework to investigate epistemological beliefs of chemistry teachers and Hsu et al. (2017) have surveyed teachers' beliefs about game-based learning with TPACK. Hughes et al. (2020) have examined teachers' pedagogical reasoning for and the technological knowledge underlying their most-valued technology-supported activities. However, because TPACK focuses on teacher understanding, it cannot be immediately and directly and generally extended to the evaluation of technology use with students in schools and higher education institutions.

Numerous frameworks can be used to evaluate the results of learning technology usage along a single dimension. For instance, the Mental Effort Scale (Paas, 1992) measures the cognitive load of learners, with Castro-Alonso et al. (2015) for example using the scale to measure the mental effort of learners completing animated Lego tasks. The Structure of the Observed Learning Outcome (SOLO) taxonomy (Biggs & Collis, 1982) evaluates conceptual knowledge (learning-related items). Padiotis and Mikropoulos (2010) have applied SOLO taxonomy to understand science knowledge of secondary school students in an educational virtual environment. Fang and Tajvidi (2018) have investigated the effects of computer simulation and animation on cognitive processes of the students in an undergraduate engineering course by using the Bloom's taxonomy. Motivation has been measured by Motivated Strategies for Learning Questionnaire (Pintrich et al., 1991) and Hsia et al. (2016) have examined the motivation of students using online peer-feedback in a dance course. Presence/community has been measured using the Presence Questionnaire (PQ). Grassini et al. (2020) use the Presence questionnaire to examine the sense of presence by using the virtual reality for training professional skills, whereas Tüzün and Özdinç (2016) have measured users' perception of presence in a 3D virtual environment using the Presence Questionnaire. However, these instruments typically serve to evaluate a single aspect of technology usage, without consideration of other important aspects of education. The Community of Inquiry (CoI) Framework (Garrison et al., 2000) does address several aspects of learning through the use of technology, but does not directly focus on investigating user's behaviour or the institutional environment, or indeed the quality of the technology itself.

Consequently, it appears that there is no broad commonly used theoretical framework for evaluating the use of technology in education across a comprehensive set of dimensions.

2.2 | Previous work on the dimensions of educational technology usage

The aspects of technology use in education that are typically evaluated in empirical studies provide another useful and important referent for considering what we should evaluate when using technology in education. A recent systematic literature review conducted by Lai and Bower (2019) examined different themes when evaluating technology use in the educational context, and identified eight dimensions—*learning outcomes*, *affective elements*, *behaviour*, *design*, *technology*, *teaching/pedagogy*, *presence/community* and *institutional environment*. Among the eight themes, *learning outcomes* was regarded as the most popular dimension when scholars evaluate the educational use of technology for learning (see, for instance, Hwang & Tsai, 2011), including evaluation of knowledge, achievement, performance, or skills development like interpersonal skills, communication skills, or motor skills (Mekonen & Fitiavana, 2021). *Affective elements* relate to learners' perceptions, intentions, preferences, attitudes, values or beliefs (e.g., see Rap & Blonder, 2017). *Behavioural aspects* consist of participation, interaction, collaboration and cooperation between or among learners (for instance, Zhang, Liu et al., 2017). *Design*, comprises course quality, course content, course structure, resources or overall design (see Sun, 2016). Another dimension, *technology*, is usually measured by its functionality, perceived usefulness, perceived ease of use, or accessibility (e.g., Han & Shin, 2016). *Teaching/pedagogy* includes pedagogical practice, teaching strategies or teaching quality (see, for instance, Martin et al., 2019). *Presence* consists of social presence, co-presence or community as well as the presence in the environment (see, e.g., Yang, 2016). Lastly, *institutional* *environment* takes into account the institutional capacity, institutional intervention, policy and support in facilitating the use of technology in teaching and learning (e.g., Snodgrass et al., 2016; Tømte et al., 2019). However, it is not known whether the constructs that are evaluated with relation to technology use in education directly reflect the values of specialists in the area of educational technology research, or whether there are mismatches between what is evaluated and the areas of perceived importance (for instance, due to convenience or expedience).

Identifying gaps between what is evaluated and what should be evaluated could result in valuable recommendations about the focus of future educational technology studies and the methods used to conduct evaluations. Therefore, this research attempts to ascertain from the perspective of field specialists, the dimensions and elements of learning technology evaluation that are of most importance, and compares their perceptions to the constructs that have been investigated in the learning technology research.

While the empirical practice with relation to the evaluation of technology use in education was readily available to the research team from their previous review (as summarized above), the values and beliefs regarding educational technology evaluation were not. To gauge the values and beliefs, a cohort of educational technology field specialists were surveyed to ascertain areas they believe are of greatest relevance of a number of questions relating to educational technology evaluation. These ratings, along with open-ended feedback about the question items, also provided a means to examine the face and content validity of a generalized instrument for evaluating the use of technology in education.

2.3 | Face and content validity in educational technology evaluation

Face validity is defined as the degree that respondents or users judge that the items of an assessment instrument are appropriate to the targeted construct and assessment objectives (Mårtensson et al., 2019; Song & Herman, 2010). Content validity refers to the extent to which the items on a measure are fairly representative of the entire domain the test seeks to measure (DeVellis, 2017; Salkind, 2010). To establish the face and content validity of items, experts are usually presented with individual items and asked to evaluate the degree to which items are relevant or representative to the construct's conceptual definition (Zaichkowsky, 1985). When researchers employ Zaichkowsky's method, in many cases items are deleted which evaluated by experts as being not representative or as a poor indicator of the construct (Hardesty & Bearden, 2004).

Prior studies about educational technologies have examined face or content validity when developing new measurement items, however the frequency and extent of such validation techniques are often limited. We re-examined the 365 empirical studies from our previous systematic review (Lai & Bower, 2019), to see which had applied face and content validity processes. Of the 365 empirical studies, 122 studies developed their own instruments, and of these, the face or content validity of instruments developed was only explicitly examined in 25 papers (see Appendix A). On average, the validity process involved five experts/specialists who knew the content and area well, with experts invited to review the instrument and to modify and reword any items which were considered inappropriate. More details of the face and content validity of the empirical studies are listed in Appendix A. This tendency for small sample face and content validity has continued more recently, with less than 10 experts typically involved in item generation and validation processes (e.g., Alharthi, 2020; binti Daud, 2021; Iqbal & Bhatti, 2020; van der Spoel et al., 2020), or occasionally involving around 10 validators (e.g., Azizi & Khatony, 2019). To the extent that this paper examines the face and content validity of instrument items based on the perceptions of a large sample of specialist educational researchers ($n = 48$), it constitutes one of the most rigorous validation processes undertaken in the educational technology field.

3 | METHODS

Domain specialists were surveyed to determine their perceptions of the relevance of different constructs relating to the evaluation of technology use. The eight dimensions identified from our previous study were operationalized into 39 Likert-scale items in the item generation phase. Those items were largely drawn from the sub-areas in the eight dimensions that emerged from the previous study, but also by referring to literature more broadly for relevant elements. The proposed items were revised in three rounds by the four members of the research team so as to improve the initial face validity of the items. This type of validity process can provide information on the representativeness and clarity of each proposed item (Mousavi et al., 2020; Rubio et al., 2003). The 39-item instrument that was presented to the field specialists is provided in Table 3 in Section 4.

3.1 | Procedures

In June 2019, 515 email invitations were sent to the authors from the original review of 365 papers, and reminders were sent 2 and 4 weeks after the first invitation. Having previously published in *Computers & Education* (2020 Impact Factor 8.538), the authors of those papers were regarded as field specialists who were familiar with the topics relating to the evaluation of technology in educational contexts. The purpose of the study was explained in the invitation email and the link to the questionnaire created using the Qualtrics online survey platform was also provided. At the very beginning of the survey, the field specialists were asked to state the dimensions they considered as most important when evaluating the use of technology in education. This question was on a separate page which could not be revisited by respondents, so that the items of technology evaluation that they identified in the first section of the survey were not affected by the 39 items they were shown later in the questionnaire. The field specialists were then asked to rate the appropriateness of the 39 proposed items for evaluating the use of technology in education on a 5-point scale from Highly irrelevant (0), Irrelevant (1), Neutral (2), Relevant (3) to Highly relevant (4). We also asked the specialists for suggestions in terms of wording, additional questions, questions to delete, other explanations and reflections related to the different dimensions of educational technology evaluation. Ethics approval was obtained for this study, and respondents received a $20 gift card as compensation for their time to participate in this study if they provided their names and email addresses for contact purposes.

Data were exported from Qualtrics into both QSR NVivo, Version 12 and Microsoft Excel 2010 for analysis purposes. First of all, the participant preferences about important dimensions of technology evaluation in open-ended responses were thematically analysed with NVivo. Researchers analysed the data sentence by sentence for each respondent in order to identify the themes (key ideas) consistently throughout the process of determining the emerging themes (Braun & Clarke, 2006), with an inductive approach employed when coding the themes (Patton, 2014; Williams & Moser, 2019). Many of the themes were relatively straight-forward to derive directly from the results, for instance, relating to technology ('it is important to consider whether used software (or hardware) is user friendly for students'), learning outcomes ('The things to be measured would, of course, be student learning, in an online learning environment'), or affective elements ('student satisfaction matter as well'), etc. The inductive approach meant that themes that hand not been identified in the previous analysis of empirical research did emerge (for instance, cost, data security). The themes that the field specialists initially identified were then compared to those that had emerged from the empirical literature for commonalities and differences in emphasis.

In addition, the field specialist relevance ratings of the individual items were analysed in the Excel spreadsheet using descriptive statistics to facilitate the amalgamation of items into factors and to enable comparison. To enable the relevance of individual items to be gauged two indices were created—a relevance index and an irrelevance index. The relevance index was calculated as the percentage of participants who rated an item as highly irrelevant or irrelevant, and the relevance index was calculated as the percentage of participants who rated an item as relevant or highly relevant. In previous research items with greater than 20% irrelevance index were considered as likely candidates for deletion from survey instruments, in alignment with Kumar and Ratnakar (2011).

As a post-hoc process, the study assimilated the feedback from the specialists to improve the content relevance and wording of the instrument items, resulting in items being reworded for clarity and better alignment with the underlying dimension of evaluation. The nature of field specialist comments, and specific responses to their recommendations, are provided in Section 4.

TABLE 1 Experience of the field specialists

Experience	Frequencies
Conduct research and develop technology in educational contexts	32
Teaching in higher education	23
Professional in the IT field	8
Teach technology-related courses	9
Teaching in schools (K-12)	6

3.2 | Participants

There was a total of 48 international educational technology research specialists who responded to the survey invitation. Their years of experience ranged from 3 to 40 years, with an average of 15 years of experience in the field. These specialists were from different areas such as educational technology designers and researchers ($N = 32$), teaching staff in higher education ($N = 23$) as well as school education ($N = 6$), noting that participants could have more than one type of field experience (Table 1).

4 | RESULTS

4.1 | Perceptions of field specialists about what should be evaluated

The responses of the specialists to the initial question about what should be evaluated when examining educational technology use included all of the eight dimensions that emerged from the systematic literature review, that is, *learning outcomes*, *affective elements*,

TABLE 2 Dimensions to consider when evaluating the use of technology

Dimensions	Frequency[a]	Percentage (%)[a]
Learning outcomes[b]	42	88
Technological aspects[b]	35	73
Affective elements[b]	18	38
Pedagogical approaches/Teaching purposes[b]	12	25
Demographic factors/Individual differences	11	23
Institutional support[b]	11	23
Behavior[b]	7	15
Design[b]	5	10
Implementation costs	4	8
Data security/privacy	3	6
Different contexts	2	4
Presence/community[b]	1	2
Others (general), e.g., will not cover all aspects	3	6

[a]Specialists could provide more than one dimension which they regarded as important in evaluating learning technology.
[b]Themes that directly correspond to the original eight themes identified in the framework.

b for in the list. others are extra ones suggested by experts *behaviour*, *design*, *technology*, *teaching/pedagogy*, *presence/community* and *institutional environment* (Table 2).

Most of the specialists (88%) believed that *learning outcomes* should be considered when evaluating the use of technology. For instance, one specialist mentioned that 'The most important dimension is learning, i.e., the increase of procedural and conceptual knowledge and skills'. The next most frequently identified evaluation term was *technological aspects* (73%), with another specialist for instance stating that 'Functionality is the most important dimension because technologies change but the basic functions technology serve remain the same (or highly similar)'. Less frequently mentioned were *affective elements* (38%), *pedagogical elements/teaching purposes* (25%), *institutional support* (23%), *behaviour* (15%), *design* (10%) and *presence/community* (2%).

The specialists also proposed other facets that could be included when evaluating educational technology application in learning and teaching. For instance, implementation costs and data security/privacy were raised by 4 and 3 out of the 48 specialists, respectively, as interesting additional considerations when evaluating educational technology usage. However, often these elements do not directly relate to the sorts of issues that educators and researchers would typically want to gauge as a result of applying learning designs with students, and instead relate more to institutional product securement processes. They were also only raised by a minority of participants. So while these were seen as important aspects of educational technology evaluation, they were not considered as crucial elements that would be typically used in surveys with students evaluating the use of technology in education. The infrequent evaluation of these aspects of technology within educational research contexts is further validated by their absence from the previous systematic analysis (Lai & Bower, 2019). Furthermore, the field specialists noted that different demographic and contextual items are important background information to help understand how contexts affected the use of technology in education.

4.2 | Perceived relevance of the eight dimensions of learning technology use

Regarding the relevance of the eight dimensions in aggregate, the descriptive statistics showed that the mean scores of item ratings across the eight dimensions ranged from 2.86/4 to 3.24/4, with technology the most relevant dimension whereas institutional environment was seen as the least relevant (Figure 1). The fact that six items had a rating of 3 or more (between 3.00 and 3.24) and the other two items had a rating of 2.96 and 2.86 (much closer to relevant than neutral) indicated strong support from the specialists with relation to the relevance of the dimensions. As well, it is apparent that there was considerable consistency across the eight dimensions in terms of expert rating of their importance, with the average ratings of all dimensions lying within 0.48 of one another (2.86–3.24). Figure 1 also indicates the variance within the eight constructs in the form of 95% confident intervals. There was the most variation in *institutional environment* relevance ratings and least variation in *technology* relevance

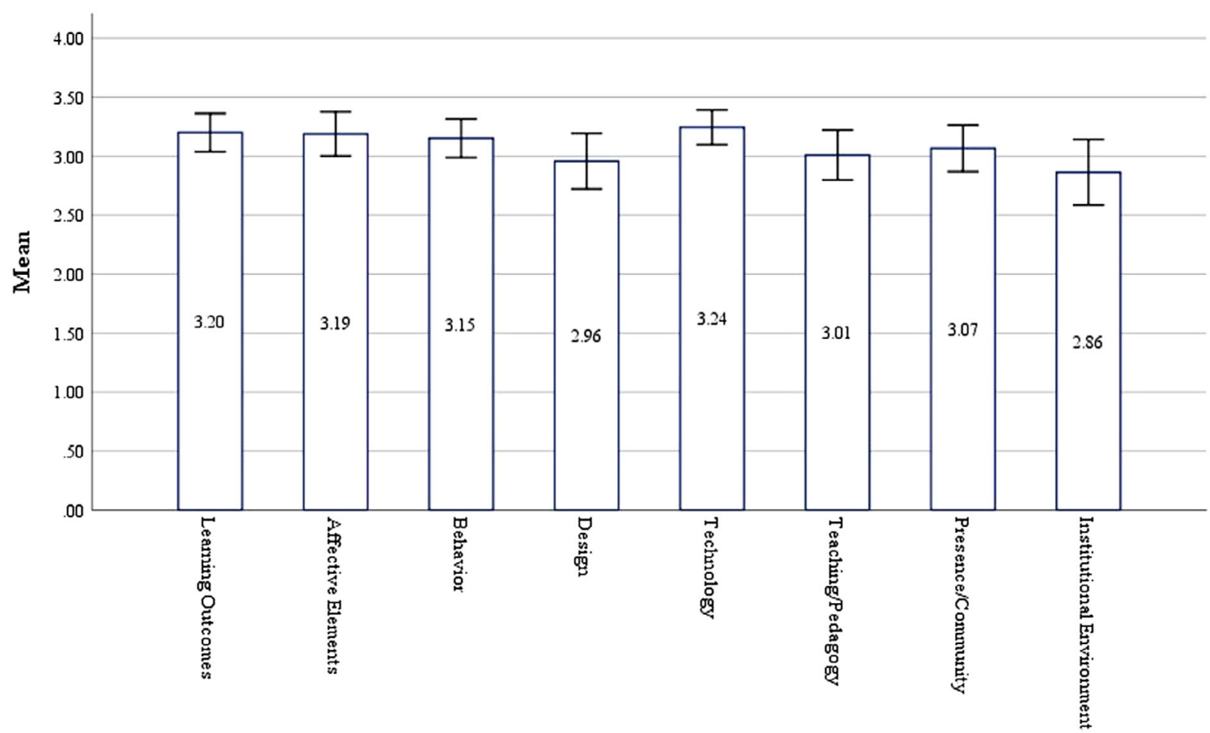

FIGURE 1 Mean relevance ratings from highly irrelevant (0) to highly relevant (4) for all items in each of the eight dimensions of the survey instrument, including 95% confidence intervals

Ask specialists whether is it valid to ask participants the questions in list. Being showing these eight, also ask in advance what should be evaluated.

TABLE 3 Sub-area results of eight dimensions

	Mean	SD	HI	I	N	R	HR	% of HI & I	% of R & HR
Dimension 1: Learning Outcomes	3.20	0.56							
Using technology increased the amount I learnt in the subject.	3.31	0.80	0	2	4	19	23	4%	88%
Using technology has led to a better learning performance in the subject.	3.35	0.76	0	2	2	21	23	4%	92%
Using technology decreased the amount of mental effort required to learn the subject matter.	2.75	1.00	1	5	10	21	11	13%	67%
Using technology increased my skills in the subject.	3.33	0.75	0	1	5	19	23	2%	88%
Using technology has improved my academic knowledge in the subject.	3.25	0.76	0	1	6	21	20	2%	85%
Dimension 2: Affective Elements	3.19	0.65							
Using technology enhanced my attitudes, values and beliefs towards the subject.	3.02	0.91	0	4	7	20	16	8%	75%
Using technology enhanced my engagement in the subject.	3.33	0.81	0	2	4	18	24	4%	88%
The use of technology helped to reduce my anxiety towards the subject.	3.19	0.84	0	2	7	19	20	4%	81%
Using technology for learning helped to improve my confidence in the subject.	3.21	0.87	1	0	8	17	21	2%	79%
Using technology enhanced my attitudes, values and beliefs towards the subject.									
Dimension 3: Behaviour	3.15	0.56							
The use of technology increased my participation in the subject.	3.25	0.72	0	1	4	25	18	2%	90%
For this topic, I was frequently using technology for my learning in the subject.	2.85	0.92	2	0	12	23	10	4%	69%
The use of technology increased the amount I could interact with others in the subject.	3.23	0.72	0	0	8	21	19	0%	83%
The use of technology increased my ability to reflect upon and regulate my learning in the subject.	3.30	0.92	1	1	6	14	25	4%	81%
Dimension 4: Design	2.96	0.81							
The use of technology resulted in a better-quality subject.	2.79	1.09	1	6	10	16	15	15%	65%
The use of technology enhanced the overall design of the subject.	2.87	0.94	1	4	6	25	11	10%	75%
The use of technology enhanced the subject content.	2.94	1.00	1	3	10	17	16	8%	69%
The use of technology enhanced the subject structure.	2.98	0.91	1	2	8	22	14	6%	75%
Dimension 5: Technology	3.24	0.51							
The technology used for learning in the subject was of high quality.	2.98	0.84	1	0	11	21	13	2%	71%
The functionality of the technology helped me to learn the subject.	3.15	0.90	1	1	7	18	19	4%	77%
The technology used for learning in the subject was easy to use.	3.40	0.67	0	0	5	18	24	0%	88%
The technology used for learning in the subject was reliable.	3.34	0.63	0	0	4	23	20	0%	90%
The technology used for learning in the subject was accessible.	3.35	0.63	0	0	4	22	20	0%	88%
Dimension 6: Teaching/Pedagogy	3.01	0.73							
The use of technology increased my overall perceptions of the quality of teaching in the subject.	2.96	0.90	2	0	8	25	12	4%	77%
The use of technology enhanced the teaching practices used in the subject.	3.09	0.90	1	1	8	20	17	4%	77%

Caution: background of specialists generally lack teaching experience (only 6).

(Continues)

TABLE 3 (Continued)

	Mean	SD	HI	I	N	R	HR	% of HI & I	% of R & HR
The use of technology improved feedback processes in the subject.	3.26	0.81	1	0	5	21	20	2%	85%
The use of technology enhanced the teaching credibility in the subject.	2.74	0.86	1	1	16	19	9	4%	58%
Dimension 7: Presence/community	3.07	0.68							
The use of technology increased my sense of being present in the class in the subject.	3.07	0.73	0	0	11	21	14	0%	73%
Using technology enhanced my sense of connection with the teacher in the subject.	3.17	0.78	0	1	8	20	18	2%	79%
Using technology enhanced my sense of connection with other students in the subject.	3.28	0.82	0	1	8	15	23	2%	79%
Using technology provided me with a stronger sense of social connection with my class in terms of being a part of a community in the subject.	3.04	0.87	0	2	11	17	17	4%	71%
The use of technology increased my sense of being present in the learning space in the subject.	2.78	0.94	1	3	12	18	11	8%	60%
Dimension 8: Institutional Environment	2.86	0.96							
The institutional support that was provided for the use of technology increased my perception of the subject.	2.64	1.13	3	3	14	12	13	13%	52%
The institutional support that was provided for the use of technology positively contributed to my experience.	2.77	1.04	2	3	11	19	12	10%	65%
The institution encourages the use of technology in education.	2.98	1.04	2	3	5	21	16	10%	77%
The institution provides the necessary infrastructure to facilitate the use of technology in education.	3.06	1.08	2	3	5	17	20	10%	77%

Abbreviations: HI, highly irrelevant; HR, highly relevant; I, irrelevant; N, neutral; R, relevant; SD, standard deviation.

ratings, indicating that experts had relatively less consensus regarding the applicability of *institutional environment* as an evaluation dimension.

The analysis of the sub-areas of each dimension provided further insights into the relevance of the dimension (see Table 3). As previously indicated, experts considered *technology* ($M = 3.24$) the most relevant dimension to consider when evaluating the use of technology in education, with the ease of technology use sub-area having the highest rating of any sub-area ($M = 3.40$). However, there was also strong support for the relevance of the other four items in the technology dimension, with the field specialists rating them between 2.98 and 3.35.

Learning outcomes was rated as the second most relevant dimension overall (after technology). The sub-area related to technology leading to a better learning performance had the highest average relevance rating ($M = 3.35$), whereas the sub-area relating to technology decreasing the amount of mental effort required was rated as least relevant with $M = 2.75$, but 67% of field specialists still thought this was relevant or highly relevant for evaluating technology use in education. The third most relevant dimension was *affective elements*, with strong support for all sub-areas demonstrated by average relevance ratings greater than 3.

Behaviour was regarded as the fourth most relevant dimension by the specialists. There was high perceived relevance of sub-areas relating to reflection ($M = 3.30$), participation ($M = 3.25$) and interaction ($M = 3.23$), though merely using technology frequently was not considered as relevant to evaluate ($M = 2.85$). *Presence/community* was rated the fifth most relevant dimension, with all items apart from being present in the learning environment receiving a relevance score greater than 3. The ratings of most items in this dimension were surprisingly high considering that only one specialist nominated this area in the opening question, and that presence is infrequently evaluated in the literature. The *teaching/pedagogy* dimension was rated after presence, with the most relevant sub-area relating to technology improving feedback processes ($M = 3.26$) and the least relating to enhancing teacher credibility ($M = 2.74$).

Design was rated as the second least relevant dimension. The sub-area about technology enhancing design of the assessments had the highest rating ($M = 3.21$), with all other sub-areas relating to design receiving a relevance score of between 2.79 and 2.98. *Institutional environment* was regarded as the least relevant dimension. However, 77% of the specialists indicated that two sub-areas in particular, institution providing necessary infrastructure ($M = 3.06$) and institution encouraging the use of technology in education ($M = 2.98$), were still relevant to evaluating the use of technology in education. Noting that the highest irrelevance score across any of the items was 15%, it appears that the field specialists agreed that all eight dimensions and their sub-areas were somewhat relevant to the evaluation of technology use in education.

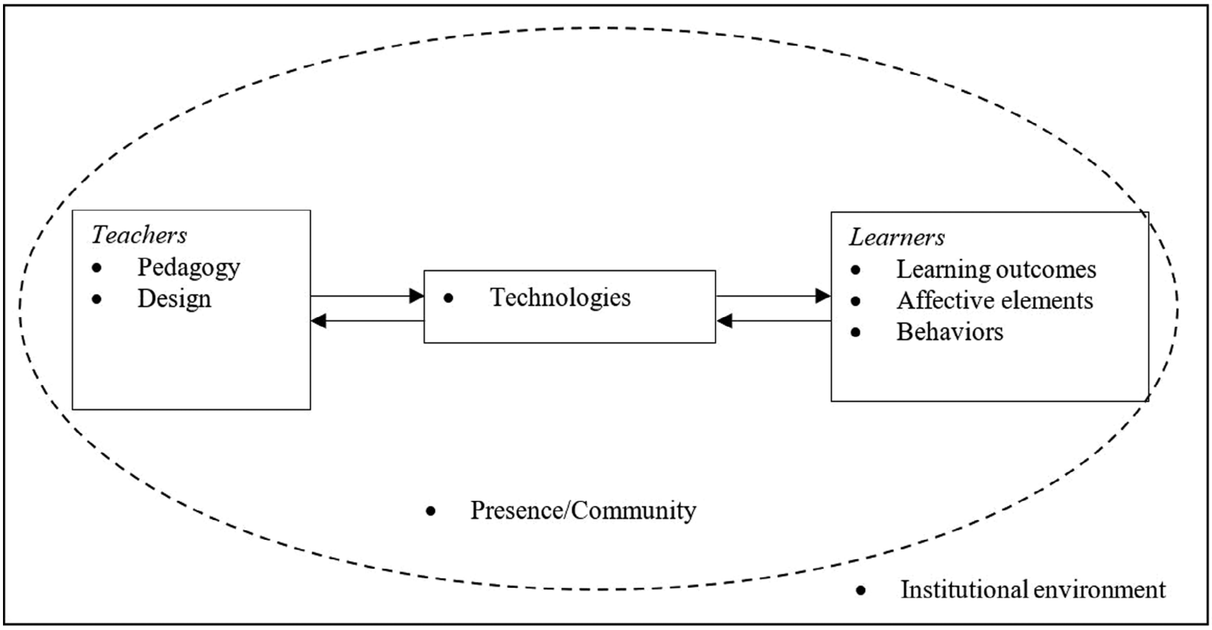

FIGURE 2 The interrelationships among the eight dimensions in this study (aligning with Bower, 2019)

4.3 | Validity of learning technology evaluation items

In addition to rating the relevance of each evaluation item, all specialists made recommendations on how to improve the items. Specialists provided several suggestions about how to improve the instrument along all dimensions, with the most suggestions for *technology* (35 suggestions), *behaviour* (32 suggestions), *teaching/pedagogy* (30 suggestions), *presence/community* (29 suggestions), *learning outcomes* (26 suggestions), *institutional environment* (25 comments), *affective elements* (24 comments) as well as *design* (21 comments).

Their feedback related to three main facets: (i) fidelity to the given dimension, (ii) the wording of items, and (iii) clarity of meanings. For example, in terms of fidelity, one specialist argued that engagement seemed to be more conative (behavioural) in nature. Another specialist thought that in the technology aspect, measuring the usefulness of the technology was essential.

Regarding wordings of items, for instance, we deleted the word 'academic' when evaluating the learning outcomes as specialists regarded it as redundant. Some specialists cast doubt on terms such as 'better-quality subject' as being unclear and subjective in meaning. Also, specialists were concerned that there was some jargon in some of the items, and the use of simpler language in some cases would be beneficial. For instance, most learners were not going to know the meaning of 'teaching credibility'. With regard to the presence and community dimension, one specialist believed that being present in the learning space could be modified to refer to immersion. We then changed the question to 'The way technology was used in this course enabled me to immerse myself in the learning environment'.

In this study, there was no item where more than 15% of specialists rated the item as irrelevant, so according to the benchmark of 20% used in previous studies (e.g., Kumar & Ratnakar, 2011), there was no cause to delete any survey items. However, according to the relevance index that we created, the item 'institutional support was provided for the use of technology increased my perception of the subject' was only rated as relevant by just over half (52%) of field specialists. As such, it could be considered as a candidate for deletion in any future application of the survey instrument to evaluate the use of technology in education.

5 | DISCUSSION

Evaluation of technology in education is a complex pursuit, with many different aspects of technology use variously evaluated in educational contexts, and different people inevitably holding a wide range of values about what should be evaluated. Findings that emerged from surveying the 48 field specialists indicated a large degree of alignment between what they felt were the important aspects of learning technology evaluation and the aspects of educational technology use that had been empirically evaluated according to the earlier systematic review. In other words, the eight dimensions identified from the previous review of what had been evaluated (Lai & Bower, 2019), namely *learning outcomes*, *affective elements*, *behaviour*, *design*, *technology*, *teaching/pedagogy*, *presence* and *institutional environment* were all generally found to be valuable by the specialists. The generally high average relevance scores for all the dimensions of *technology* ($M = 3.24/4$), *learning outcomes* ($M = 3.20/4$), *affective elements* ($M = 3.19/4$), *behaviour* ($M = 3.15/4$), *teaching/pedagogy* ($M = 3.01/4$), *presence* ($M = 3.07/4$), *design* ($M = 2.96/4$), and *institutional environment* ($M = 2.86/4$), indicate a reasonably high degree of perceived relevance for each construct. Interestingly, previous research in this domain has shown that evaluation of technology use in education is generally limited to focusing on one, two or three dimensions, with rarely five but never all eight dimensions (Lai & Bower, 2019). This is the first study to propose evaluation using all eight dimensions, and surprisingly there was a general consensus among specialists that all eight dimensions were relevant. This would appear to imply that there is scope for educational researchers to investigate more aspects of learning technology usage in their evaluations than are currently being examined.

The various dimensions that have been empirically investigated can be related to one another as shown in Figure 2. The dimensions *learning outcomes*, *affective elements*, as well as *behaviour* all constitute aspects of learners, highlighting how evaluation of elements such as knowledge acquisition, attitudes and interaction between students is critical. Teachers on the other hand embrace technology to improve the overall *design* of tasks and platforms, as well as the *pedagogies* that are used. In addtion, there are circumscriptive influences relating to the use of technology in education. Most immediately, the sense of *presence and community* that is experienced will influence the quality of experience. More broadly, the *institutional environment* plays a higher level role in influencing technology use in education.

From the literature review, it is apparent that none of the previous models/frameworks we mentioned previously, that is, Activity Theory, the SOLO Taxonomy, Mental Effort Scale, Motivated Strategies for Learning Questionnaire, TPACK, TAM, UTUAT, UTUAT2, CoI, and Institutional Theory, actually integrated all dimensions. That is to say, commonly used theories that are used in educational technology evaluation studies do not comprehensively account for all dimensions of educational technology use that are valued and investigated in aggregate. Also, many of the models have been borrowed from other fields and do not specifically relate to evaluating the use of technology in education. This leaves the educational technology field with a lamentable gap between theory and practice when it comes to evaluating the use of technology in education.

Interestingly, Figure 2 closely aligns with a recent *Technology-Mediated Learning Theory model* (Bower, 2019) that was derived by synthesizing relevant aspects of theories that are commonly referenced when studying educational technology usage. Thus, while none of the long-standing and common theoretical models alone appeared to capture the broad aspects involved in the evaluation of technology use in education, it appears that when a synthesis of theoretical models is compared with a comprehensive analysis of empirical evaluation practices, there is strong alignment.

The wording, semantic and applicability suggestions that arose from the specialists as part of the content and face validation process would undoubtedly improve the quality of the final instrument, highlighting the utility of rigorous validation processes when designing instruments generally. Creswell and Clark (2017) suggest that using expert validation during the development phase improves the validity of the items because researchers can understand more fully the phenomena of interest and their interpretations before attempting to validate the items associated with the situation. Previous studies involving rigorous development of a new instrument often do incorporate an expert validation process (Aesaert et al., 2014; Baran & Uygun, 2016; Iqbal & Bhatti, 2020; Janssen et al., 2013; Teo, 2013; Zhang, 2007), although rarely with the sample size in this study. Findings from the literature review indicate that only a minority of educational technology studies involving instrument development perform face and content validity checks, and even then it is typically with small samples of respondents ($n \leq 5$). The literature is diverse regarding the number of experts needed for content and face validation, with suggestions ranging between two and 20 experts (McCoach et al., 2013). To that extent, the number of experts in this study exceeded the recommended guidelines. However, drawing together 48 international specialists generated far more suggestions and instrument refinement regarding items for evaluating learning technology in educational contexts than if a smaller sample had been used. As such, we recommend that future studies involving instrument development drawing upon more extensive samples of field specialists (greater than 25) if possible for the purpose of face and content validation.

Specifically, the feedback from field specialists in this study provided detailed information that validated and also improved our a comprehensive instrument for evaluating the use of technology in education. To check the content relevance and wording, any ambiguous or unclear items were reworded. For instance, it was noted that all items were phrased positively, and that some negative items that required reversed coding should be added to avoid responder directionality bias (Grover & Vriens, 2006).

The breadth of items that were identified as relevant by the field specialists indicates the complex and multifaceted nature of evaluating educational technology use. To finalize the scale and confirm its dimensionality, reliability and validity, a large quantitative study has been undertaken (Lai, De Nobile, Bower & Breyer, in press). Principal Component Analysis and Confirmatory Factor Analysis were used to test the overall construct validity (Canivez et al., 2019; Wallace, 2004), resulting in a 28 item validated instrument (see Lai, De Nobile, Bower & Breyer, in press). Further studies could also focus on creating an instrument to be completed by stakeholders apart from students, for instance, educators, executives, technology developers. These could include aspects such as security, privacy and cost.

While this paper constitutes one of the largest validation studies conducted in the educational technology field, it should be noted that the findings of the present study still need to be interpreted carefully because consultation of different or more field specialists may have led to different results. Similar studies in the future could include a wider variety of respondents and consider different cohorts.

6 | CONCLUSION

Given the increasing use of technology in education, the contribution of our study in is four-fold: (1) understanding the values of education specialists with relation to evaluating educational technology use; (2) exposing the gap between practice and values with relation to the evaluation of technology use in education, (3) highlighting the lack of validation that has previously occurred in the evaluation of educational technology use, and (4) showcasing the efficacy of an instrument incorporating eight distinctive dimensions that align with the values and practices of educational technology researchers. This work is particularly important in light of the relative paucity of critical studies into the evaluation of technology use in education.

Previous research in this domain has shown that evaluation of technology use in education is generally limited to focusing on one, two or three dimensions, and this is the first study to include all eight dimensions. Unexpectedly, there was general consensus between the field specialists that all eight dimensions that we were identified in the literature review were of value, from *technology* ($M = 3.24/4$), *learning outcomes* ($M = 3.20/4$), *affective elements* ($M = 3.19/4$), *behaviour* ($M = 3.15/4$), *presence/community* ($M = 3.07/4$), *teaching/pedagogy* ($M = 3.01/4$) which all had mean relevance ratings of greater than 3 (between relevant and highly relevant). While *design* ($M = 2.96/4$) and *institutional environment* ($M = 2.86/4$) had slightly lower ratings, they were still deemed much closer to relevant than neutral.

In this study, the creation of the items followed a structured procedure that involved first identifying the critical dimensions from a systematic literature review conducted by the authors. Subsequently, measurement items were composed among the sub-areas and those items were further evaluated by the four researchers in three rounds to uphold the initial face validity of the measurement items. In addition, in order to make robust and vigorous decisions about which item to keep and which item to delete, we calculated the 'not relevant' percentages and 'relevant' percentages per each item. The 'not relevant' category was calculated as the percentage of specialists that selected irrelevant or highly irrelevant for a particular item, whereas relevant category was the percentages of specialists picked relevant or highly relevant for that question. Language and wording of the items were amended according to the comments from the specialists.

Understanding the process of conducting face and content validity is important for researchers and practitioners/educators. The methodological limitations in many previous studies that analyse the impact of learning technologies have been noted, such as the sample size issues, the construct and content validity of the measurement items, as well as the generalizability of the results (Cox & Marshall, 2007; Kim & Park, 2019; Kirkwood & Price, 2015). Researchers in educational field are obliged to critically review the measurements they use in their studies, including broadening out the number of dimensions that are examined. If researchers plan to develop or revise measures, it is advisable that they follow established guidelines. With the collective agreement of the experts and researchers on the validity of the measurement items, we can be more confident that scales are measuring what they are expected to measure. Also, the panel of field specialists indicated that they were satisfied with the relevance of the items and hence we kept all eight dimensions in our instrument in accordance with Sreejesh et al. (2014). Overall, the face and content validity study provided valuable direction for the revision of the instrument.

This paper aimed to identify relevant evaluation items using a thorough, rigorous, robust, and multifaceted evaluation approach relating to the use of technology in education. Among the educational technology research literature, there does not appear to be a comprehensive instrument that is empirically grounded and has been developed to systematically evaluate the use of technology from the learner's perspective. Reliable comparison of the use of technologies across contexts using a generalized instrument would mean that educators can detect which approaches are effective across implementations, which in turn can inform the design of technology-enhanced curriculum and guidelines for the development and use of technology in education. The development of a robust and transferable evaluation instrument could help educators in both the school and Higher Education sectors more consistently evaluate and implement technology to improve overall student learning and experiences. Subsequent work confirming the factor structure and deriving a 28 item validated instrument using a large sample of students (Lai, De Nobile, Bower & Breyer, in press) serves to reinforce the overall veracity of ratings and perceptions expressed by the field specialists in this study.

ACKNOWLEDGEMENTS

This research was approved by the Macquarie University Human Research Ethics Committee (approval ref.: 5201952329337).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12645.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Jennifer W. M. Lai https://orcid.org/0000-0001-9042-7064
Matt Bower https://orcid.org/0000-0002-4161-5816
John De Nobile https://orcid.org/0000-0002-5640-611X
Yvonne Breyer https://orcid.org/0000-0003-3936-5050

REFERENCES

Aesaert, K., van Nijlen, D., Vanderlinde, R., & van Braak, J. (2014). Direct measures of digital information processing and communication skills in primary education: Using item response theory for the development and validation of an ICT competence scale. *Computers & Education*, *76*, 168–181. https://doi.org/10.1016/j.compedu.2014.03.013

Alharthi, M. (2020). Students' attitudes toward the use of technology in online courses. *International Journal of Technology in Education*, *3*(1), 14–23.

Alqahtani, A. Y., & Rajkhan, A. A. (2020). E-learning critical success factors during the COVID-19 pandemic: A comprehensive analysis of e-learning managerial perspectives. *Education Sciences*, *10*(9), 216. https://doi.org/10.3390/educsci10090216

Arpaci, I. (2017). Antecedents and consequences of cloud computing adoption in education to achieve knowledge management. *Computers in Human Behavior*, *70*, 382–390. https://doi.org/10.1016/j.chb.2017.01.024

Ashrafzadeh, A., & Sayadian, S. (2015). University instructors' concerns and perceptions of technology integration. *Computers in Human Behavior*, *49*, 62–73. https://doi.org/10.1016/j.chb.2015.01.071

Azizi, S. M., & Khatony, A. (2019). Investigating factors affecting on medical sciences students' intention to adopt mobile learning. *BMC Medical Education*, *19*(1), 1–10. https://doi.org/10.1186/s12909-019-1831-4

Baran, E., & Uygun, E. (2016). Putting technological, pedagogical, and content knowledge (TPACK) in action: An integrated TPACK-design-based learning (DBL) approach. *Australasian Journal of Educational Technology*, *32*(2), 47–63. https://doi.org/10.14742/ajet.2551

Baser, D., Kopcha, T. J., & Ozden, M. Y. (2016). Developing a technological pedagogical content knowledge (TPACK) assessment for preservice teachers learning to teach English as a foreign language. *Computer Assisted Language Learning*, *29*(4), 749–764. https://doi.org/10.1080/09588221.2015.1047456

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. Academic Press.

binti Daud, R. (2021). Face and content validity for the special education leadership (integration) questionnaire in Malaysia. *Turkish Journal of Computer and Mathematics Education*, *12*(11), 5172–5178.

Bowen, B. D., DeLuca, V. W., & Franzen, M. M. S. (2016). Measuring how the degree of content knowledge determines performance outcomes in an engineering design-based simulation environment for middle school students. *Computers & Education*, *92–93*, 117–124. https://doi.org/10.1016/j.compedu.2015.10.005

Bower, M. (2019). Technology-mediated learning theory. *British Journal of Educational Technology*, *50*(3), 1035–1048. https://doi.org/10.1111/bjet.12771

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Briz-Ponce, L., Pereira, A., Carvalho, L., Juanes-Méndez, J. A., & García-Peñalvo, F. J. (2017). Learning with mobile technologies—Students' behavior. *Computers in Human Behavior*, *72*, 612–620. https://doi.org/10.1016/j.chb.2016.05.027

Bulman, G., & Fairlie, R. W. (2016). Technology and education: Computers, software, and the internet. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 5, pp. 239–280). Elsevier.

Burch, P. (2010). The bigger picture: Institutional perspectives on interim assessment technologies. *Peabody Journal of Education*, *85*(2), 147–162. https://doi.org/10.1080/01619561003685288

Canivez, G. L., Watkins, M. W., & McGill, R. J. (2019). Construct validity of the Wechsler intelligence scale for children – Fifth UKedition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology*, *89*(2), 195–224. https://doi.org/10.1111/bjep.12230

Carver, L. B. (2016). Teacher perception of barriers and benefits in K-12 technology usage. *Turkish Online Journal of Educational Technology*, *15*(1), 110–116 (1303-6521).

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015). Animations showing Lego manipulative tasks: Three potential moderators of effectiveness. *Computers & Education*, *85*, 1–13. https://doi.org/10.1016/j.compedu.2014.12.022

Chen, C.-P., Lai, H.-M., & Ho, C.-Y. (2015). Why do teachers continue to use teaching blogs? The roles of perceived voluntariness and habit. *Computers & Education*, *82*, 236–249. https://doi.org/10.1016/j.compedu.2014.11.017

Cheng, M.-T., Lin, Y.-W., & She, H.-C. (2015). 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. https://doi.org/10.1016/j.compedu.2015.03.007

Chien, K.-P., Tsai, C.-Y., Chen, H.-L., Chang, W.-H., & Chen, S. (2015). Learning differences and eye fixation patterns in virtual and physical science laboratories. *Computers & Education*, *82*, 191–201. https://doi.org/10.1016/j.compedu.2014.11.023

Chuang, H.-H., Weng, C.-Y., & Huang, F.-C. (2015). A structure equation model among factors of teachers' technology integration practice and their TPCK. *Computers & Education*, *86*, 182–191. https://doi.org/10.1016/j.compedu.2015.03.016

Cox, M. J. (2013). Formal to informal learning with IT: Research challenges and issues for e-learning. *Journal of Computer Assisted Learning*, *29*(1), 85–105. https://doi.org/10.1111/j.1365-2729.2012.00483.x

Cox, M. J., & Marshall, G. (2007). Effects of ICT: Do we know what we should know? *Education and Information Technologies*, *12*(2), 59–70. https://doi.org/10.1007/s10639-007-9032-x

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. SAGE Publications.

Curtis, M. (2020). Toward understanding secondary teachers' decisions to adopt geospatial technologies: An examination of Everett Rogers' diffusion of innovation framework. *Journal of Geography*, *119*(5), 147–158. https://doi.org/10.1080/00221341.2020.1784252

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

del Barrio-Garcia, S., Arquero, J. L., & Romero-Frias, E. (2015). Personal learning environments acceptance model: The role of need for cognition, e-learning satisfaction and students' perceptions. *Educational Technology & Society*, *18*(3), 129–141.

Deng, F., Chai, C. S., So, H.-J., Qian, Y., & Chen, L. (2017). Examining the validity of the technological pedagogical content knowledge (TPACK) framework for preservice chemistry teachers. *Australasian Journal of Educational Technology*, *33*(3), 1–14 (1449-5554).

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.

Diep, N. A., Cocquyt, C., Zhu, C., & Vanwing, T. (2016). Predicting adult learners' online participation: Effects of altruism, performance expectancy, and social capital. *Computers & Education*, *101*, 84–101. https://doi.org/10.1016/j.compedu.2016.06.002

Fang, N., & Tajvidi, M. (2018). The effects of computer simulation and animation (CSA) on students' cognitive processes: A comparative case study in an undergraduate engineering course. *Journal of Computer Assisted Learning*, *34*(1), 71–83. https://doi.org/10.1111/jcal.12215

Fang, S.-C., & Hsu, Y.-S. (2017). Understanding science teachers' enactments of a computer-based inquiry curriculum. *Computers & Education*, *112*, 69–82. https://doi.org/10.1016/j.compedu.2017.05.004

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. https://doi.org/10.1016/S1096-7516(00)00016-6

González-Marcos, A., Alba-Elías, F., Navaridas-Nalda, F., & Ordieres-Meré, J. (2016). Student evaluation of a virtual experience for project management learning: An empirical study for learning improvement. *Computers & Education*, *102*, 172–187. https://doi.org/10.1016/j.compedu.2016.08.005

Grassini, S., Laumann, K., & Rasmussen Skogstad, M. (2020). The use of virtual reality alone does not promote training performance (but sense of presence does). *Frontiers in Psychology*, *11*, 1743–1743. https://doi.org/10.3389/fpsyg.2020.01743

Grover, R., & Vriens, M. (Eds.). (2006). *The handbook of marketing research: Uses, misuses, and future advances*. SAGE Publications.

Guo, Y. R., Goh, D. H.-L., Luyt, B., Sin, S.-C. J., & Ang, R. P. (2015). The effectiveness and acceptance of an affective information literacy tutorial. *Computers & Education*, *87*, 368–384. https://doi.org/10.1016/j.compedu.2015.07.015

Han, I., & Shin, W. S. (2016). The use of a mobile learning management system and academic achievement of online students. *Computers & Education*, *102*, 79–89. https://doi.org/10.1016/j.compedu.2016.07.003

Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development: Implications for improving face validity of measures of unobservable constructs. *Journal of Business Research*, *57*(2), 98–107. https://doi.org/10.1016/S0148-2963(01)00295-8

Holmgren, R., Haake, U., & Söderström, T. (2017). Firefighter learning at a distance – A longitudinal study. *Journal of Computer Assisted Learning*, *33*(5), 500–512. https://doi.org/10.1111/jcal.12196

Hsia, L.-H., Huang, I., & Hwang, G.-J. (2016). Effects of different online peer-feedback approaches on students' performance skills, motivation and self-efficacy in a dance course. *Computers & Education*, *96*, 55–71. https://doi.org/10.1016/j.compedu.2016.02.004

Hsu, C.-C., Chiu, C.-H., Lin, C.-H., & Wang, T.-I. (2015). Enhancing skill in constructing scientific explanations using a structured argumentation scaffold in scientific inquiry. *Computers & Education*, *91*, 46–59. https://doi.org/10.1016/j.compedu.2015.09.009

Hsu, C.-Y., Tsai, M.-J., Chang, Y.-H., & Liang, J.-C. (2017). Surveying in-service teachers' beliefs about game-based learning and perceptions of technological pedagogical and content knowledge of games. *Educational Technology & Society*, *20*(1), 134–143.

Huang, C.-F., Nien, W.-P., & Yeh, Y.-S. (2015). Learning effectiveness of applying automated music composition software in the high grades of elementary school. *Computers & Education*, *83*, 74–89. https://doi.org/10.1016/j.compedu.2015.01.003

Hughes, J. E., Cheah, Y. H., Shi, Y., & Hsiao, K. H. (2020). Preservice and inservice teachers' pedagogical reasoning underlying their most-valued technology-supported instructional activities. *Journal of Computer Assisted Learning*, *36*(4), 549–568. https://doi.org/10.1111/jcal.12425

Hung, S.-T. A. (2016). Enhancing feedback provision through multimodal video technology. *Computers & Education*, *98*, 90–101. https://doi.org/10.1016/j.compedu.2016.03.009

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. https://doi.org/10.1111/j.1467-8535.2011.01183.x

Inan, F., & Lowther, D. (2010). Factors affecting technology integration in K-12 classrooms: A path model. *Educational Technology Research and Development*, *58*(2), 137–154. https://doi.org/10.1007/s11423-009-9132-y

Iqbal, S., & Bhatti, Z. A. (2020). A qualitative exploration of teachers' perspective on smartphones usage in higher education in developing countries. *International Journal of Educational Technology in Higher Education*, *17*(1), 1–16. https://doi.org/10.1186/s41239-020-00203-4

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 and Development*, *64*(2), 245–262. https://doi.org/10.1007/s11423-015-9409-2

Jan, P.-T., Lu, H.-P., & Chou, T.-C. (2012). The adoption of e-learning: An institutional theory perspective. *Turkish Online Journal of Educational Technology*, *11*(3), 326–343.

Janssen, J., Stoyanov, S., Ferrari, A., Punie, Y., Pannekeet, K., & Sloep, P. (2013). Experts' views on digital competence: Commonalities and differences. *Computers & Education*, *68*, 473–481. https://doi.org/10.1016/j.compedu.2013.06.008

Kaptelinin, V., & Nardi, B. (2007). Acting with technology: Activity theory and interaction design. *First Monday*, *12*(4). 196. https://doi.org/10.5210/fm.v12i4.1772

Ke, F., & Carafano, P. (2016). Collaborative science learning in an immersive flight simulation. *Computers & Education*, *103*, 114–123. https://doi.org/10.1016/j.compedu.2016.10.003

Kim, J. H., & Park, H. (2019). Effects of smartphone-based mobile learning in nursing education: A systematic review and meta-analysis. *Asian Nursing Research*, *13*(1), 20–29. https://doi.org/10.1016/j.anr.2019.01.005

Kirkwood, A., & Price, L. (2015). Achieving improved quality and validity: Reframing research and evaluation of learning technologies. *European Journal of Open, Distance and E-learning*, *18*(1), 102–115.

Kirschner, P. A., & Kester, L. (2016). *Towards a research agenda for educational technology research. The Wiley handbook of learning technology* (pp. 523–541). Wiley Blackwell.

Kleinheksel, A. J., & Ritzhaupt, A. D. (2017). Measuring the adoption and integration of virtual patient simulations in nursing education: An exploratory factor analysis. *Computers & Education*, *108*, 11–29. https://doi.org/10.1016/j.compedu.2017.01.005

Kumar, P. G., & Ratnakar, R. (2011). A scale to measure farmers' attitude towards ICT-based extension services. *Indian Research Journal of Extension Education*, *11*(1), 109–112.

Kusuma, A. C., Basori, B., & Budiyanto, C. W. (2020). Assessing lecturers' acceptance towards online learning system: The institutional theory perspective. *Indonesian Journal of Informatics Education*, *4*(2), 62–71. https://doi.org/10.20961/ijie.v4i2.47271

Lai, J. W., & Bower, M. (2019). How is the use of technology in education evaluated? A systematic review. *Computers & Education*, *133*, 27–42. https://doi.org/10.1016/j.compedu.2019.01.010

Lai, J. W. M., De Nobile, J., Bower, M., & Breyer, Y. (in press). Comprehensive Evaluation of the Use of Technology in Education – Validation with a cohort of global open online learners. *Education & Information Technologies*.

Law, E. L.-C., & Sun, X. (2012). Evaluating user experience of adaptive digital educational games with activity theory. *International Journal of Human-Computer Studies*, *70*(7), 478–497. https://doi.org/10.1016/j.ijhcs.2012.01.007

Li, S.-C. S., & Huang, W.-C. (2016). Lifestyles, innovation attributes, and teachers' adoption of game-based learning: Comparing non-adopters with early adopters, adopters and likely adopters in Taiwan. *Computers & Education*, *96*, 29–41. https://doi.org/10.1016/j.compedu.2016.02.009

Lord, A. Y. Z., Chen, M.-P., Cheng, Y.-Y., Tai, K.-C., & Pan, W.-H. (2017). Enhancing nutrition-majored students' reflective judgment through online collective reflection. *Computers & Education*, *114*, 298–308. https://doi.org/10.1016/j.compedu.2017.07.010

Mårtensson, P., Fors, U., Fröberg, E., Zander, U., & Nilsson, G. H. (2019). Quality of research practice - an interdisciplinary face validity evaluation of a quality model. *PLoS One*, *14*(2), e0211636. https://doi.org/10.1371/journal.pone.0211636

Martin, F., Budhrani, K., Kumar, S., & Ritzhaupt, A. (2019). Award-winning faculty online teaching practices: Roles and competencies. *Online Learning Journal*, *23*(1), 184–205.

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain: School and corporate applications* (Vol. 36). Springer.

Mekonen, Y. K., & Fitiavana, R. A. (2021). Assessment of learning outcomes in higher education: Review of literature. *Assessment of Learning Outcomes in Higher Education: Review of literature*, *71*(1), 8–8.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Mousavi, A., Mohammadi, A., Mojtahedzadeh, R., Shirazi, M., & Rashidi, H. (2020). E-learning educational atmosphere measure (EEAM): A new instrument for assessing e-students' perception of educational environment. *Research in Learning Technology*, *28*, 1–12. https://doi.org/10.25304/rlt.v28.2308

Olelewe, C. J., & Agomuo, E. E. (2016). Effects of B-learning and F2F learning environments on students' achievement in QBASIC programming. *Computers & Education*, *103*, 76–86. https://doi.org/10.1016/j.compedu.2016.09.012

Oliver, M. (2000). An introduction to the evaluation of learning technology. *Educational Technology & Society*, *3*(4), 20–30. http://www.jstor.org/stable/jeductechsoci.3.4.20

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. https://doi.org/10.1037/0022-0663.84.4.429

Padiotis, I., & Mikropoulos, T. A. (2010). Using SOLO to evaluate an educational virtual environment in a technology education setting. *Educational Technology & Society*, *13*(3), 233–245.

Patton, M. Q. (2014). *Qualitative research and evaluation methods: Integrating theory and practice*. SAGE Publications.

Patwardhan, M., & Murthy, S. (2015). When does higher degree of interaction lead to higher learning in visualizations? Exploring the role of 'interactivity enriching features'. *Computers & Education*, *82*, 292–305. https://doi.org/10.1016/j.compedu.2014.11.018

Peña-Ayala, A., Sossa, H., & Méndez, I. (2014). Activity theory as a framework for building adaptive e-learning systems: A case to provide empirical evidence. *Computers in Human Behavior*, *30*, 131–145. https://doi.org/10.1016/j.chb.2013.07.057

Phillips, R., Kennedy, G., & McNaught, C. (2012a). The role of theory in learning technology evaluation research. *Australasian Journal of Educational Technology*, *28*(7), 1103–1118.

Phillips, R., McNaught, C., & Kennedy, G. (2012b). *Evaluating e-learning: Guiding research and practice*. Routledge.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. The University of Michigan.

Pirkkalainen, H., Pawlowski, J. M., & Pappa, D. (2017). Educators' open educational collaboration online: The dilemma of emotional ownership. *Computers & Education*, *106*, 119–136. https://doi.org/10.1016/j.compedu.2016.12.005

Poole, D. M., & Preciado, M. K. (2016). Touch typing instruction: Elementary teachers' beliefs and practices. *Computers & Education*, *102*, 1–14. https://doi.org/10.1016/j.compedu.2016.06.008

Porter, W. W., Graham, C. R., Bodily, R. G., & Sandberg, D. S. (2016). A qualitative analysis of institutional drivers and barriers to blended learning adoption in higher education. *The Internet and Higher Education*, *28*, 17–27. https://doi.org/10.1016/j.iheduc.2015.08.003

Porter, W. W., Graham, C. R., Spring, K. A., & Welch, K. R. (2014). Blended learning in higher education: Institutional adoption and implementation. *Computers & Education*, *75*, 185–195. https://doi.org/10.1016/j.compedu.2014.02.011

Prisacari, A. A., & Danielson, J. (2017). Rethinking testing mode: Should I offer my next chemistry test on paper or computer? *Computers & Education*, *106*, 1–12. https://doi.org/10.1016/j.compedu.2016.11.008

Rambe, P. (2012). Activity theory and technology mediated interaction: Cognitive scaffolding using question-based consultation on "Facebook". *Australasian Journal of Educational Technology*, *28*(8), 1333.

Ranellucci, J., Rosenberg, J., & Poitras, E. (2020). Exploring pre-service teachers' use of technology: The technology acceptance model and expectancy-value theory. *Journal of Computer Assisted Learning.*, *36*, 810–824. https://doi.org/10.1111/jcal.12459

Rap, S., & Blonder, R. (2017). Thou shall not try to speak in the Facebook language: Students' perspectives regarding using Facebook for chemistry learning. *Computers & Education*, *114*, 69–78. https://doi.org/10.1016/j.compedu.2017.06.014

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). The Free Press.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94–104.

Sáez-López, J.-M., Román-González, M., & Vázquez-Cano, E. (2016). Visual programming languages integrated across the curriculum in elementary school: A two year case study using "scratch" in five schools. *Computers & Education*, *97*, 129–141. https://doi.org/10.1016/j.compedu.2016.03.003

Salkind, N. J. (2010). *Encyclopedia of research design*. SAGE Publications.

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests and identities* (4th ed.). SAGE Publications.

Shroff, R. H., Deneen, C. C., & Ng, E. M. W. (2011). Analysis of the technology acceptance model in examining students' behavioural intention to use an e-portfolio system. *Australasian Journal of Educational Technology*, *27*(4), 600–618.

Snodgrass, M. R., Israel, M., & Reese, G. C. (2016). Instructional supports for students with disabilities in K-5 computing: Findings from a cross-case analysis. *Computers & Education*, *100*, 1–17. https://doi.org/10.1016/j.compedu.2016.04.011

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the what works clearinghouse (phase I). *Educational Evaluation and Policy Analysis*, *32*(3), 351–371. https://doi.org/10.3102/0162373710373389

Sreejesh, S., Mohapatra, S., & Anusree, M. R. (2014). *Business research methods: An applied orientation*. Springer International Publishing.

Sun, J. (2016). Multi-dimensional alignment between online instruction and course technology: A learner-centered perspective. *Computers & Education*, *101*, 102–114. https://doi.org/10.1016/j.compedu.2016.06.003

Tamilmani, K., Rana, N. P., & Dwivedi, Y. K. (2017). A systematic review of citations of UTAUT2 article and its usage trends. In A. K. Kar, et al. (Eds.), *Digital nations—Smart cities, innovation, and sustainability. 16th IFIP WG 6.11 conference on e-business, e-services, and e-society* (Vol. 10595, pp. 38–49). Springer.

Tarhini, A., Hone, K., & Liu, X. (2014). Measuring the moderating effect of gender and age on e-learning acceptance in England: A structural equation modeling approach for an extended technology acceptance model. *Journal of Educational Computing Research*, *51*(2), 163–184. https://doi.org/10.2190/EC.51.2.b

Teo, T. (2013). An initial development and validation of a digital natives assessment scale (DNAS). *Computers & Education*, *67*, 51–57. https://doi.org/10.1016/j.compedu.2013.02.012

Tharp, D., & Chamberlain, G. (2016). University learning technology control and security: Requires teamwork to succeed. In N. Rushby & D. Surry (Eds.), *Wiley handbook of learning technology* (pp. 348–371). John Wiley & Sons, Inc.

Tømte, C. E., Fossland, T., Aamodt, P. O., & Degn, L. (2019). Digitalisation in higher education: Mapping institutional approaches for teaching and learning. *Quality in Higher Education*, *25*(1), 98–114. https://doi.org/10.1080/13538322.2019.1603611

Tsai, F.-H., Tsai, C.-C., & Lin, K.-Y. (2015). The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. *Computers & Education*, *81*, 259–269. https://doi.org/10.1016/j.compedu.2014.10.013

Türkay, S. (2016). The effects of whiteboard animations on retention and subjective experiences when learning advanced physics topics. *Computers & Education*, *98*, 102–114. https://doi.org/10.1016/j.compedu.2016.03.004

Tüzün, H., & Özdinç, F. (2016). The effects of 3D multi-user virtual environments on freshmen university students' conceptual and spatial learning and presence in departmental orientation. *Computers & Education*, *94*, 228–240. https://doi.org/10.1016/j.compedu.2015.12.005

van der Spoel, I., Noroozi, O., Schuurink, E., & van Ginkel, S. (2020). Teachers' online teaching expectations and experiences during the Covid19-pandemic in The Netherlands. *European Journal of Teacher Education*, *43*(4), 623–638. https://doi.org/10.1080/02619768.2020.1821185

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.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178.

Wallace, C. J. (2004). Confirmatory factor analysis of the cognitive failures questionnaire: Evidence for dimensionality and construct validity. *Personality and Individual Differences*, *37*(2), 307–324. https://doi.org/10.1016/j.paid.2003.09.005

Wanner, T., & Palmer, E. (2015). Personalising learning: Exploring student and teacher perceptions about flexible learning and assessment in a flipped university course. *Computers & Education*, *88*, 354–369. https://doi.org/10.1016/j.compedu.2015.07.008

Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, *15*(1), 45–55.

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, *140*(103), 599. https://doi.org/10.1016/j.compedu.2019.103599

Yang, J., & Huang, R. (2015). Development and validation of a scale for evaluating technology-rich classroom environment. *Journal of Computers in Education*, *2*(2), 145–162. https://doi.org/10.1007/s40692-015-0029-y

Yang, S.-H. (2016). Conceptualizing effective feedback practice through an online community of inquiry. *Computers & Education*, *94*, 162–177. https://doi.org/10.1016/j.compedu.2015.10.023

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of Consumer Research*, *12*(3), 341–352.

Zhang, R.-C., Lai, H.-M., Cheng, P.-W., & Chen, C.-P. (2017a). Longitudinal effect of a computer-based graduated prompting assessment on students' academic performance. *Computers & Education*, *110*, 181–194. https://doi.org/10.1016/j.compedu.2017.03.016

Zhang, S., Liu, Q., Chen, W., Wang, Q., & Huang, Z. (2017b). Interactive networks and social knowledge construction behavioral patterns in primary school teachers' online collaborative learning activities. *Computers & Education*, *104*, 1–17. https://doi.org/10.1016/j.compedu.2016.10.011

Zhang, Y. (2007). Development and validation of an internet use attitude scale. *Computers & Education*, *49*(2), 243–253. https://doi.org/10.1016/j.compedu.2005.05.005

Zheng, X. L., Kim, H. S., Lai, W. H., & Hwang, G. J. (2020). Cognitive regulations in ICT-supported flipped classroom interactions: An activity theory perspective. *British Journal of Educational Technology*, *51*(1), 103–130. https://doi.org/10.1111/bjet.12763

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Lai, J. W. M., Bower, M., De Nobile, J., & Breyer, Y. (2022). What should we evaluate when we use technology in education? *Journal of Computer Assisted Learning*, *38*(3), 743–757. https://doi.org/10.1111/jcal.12645