Computers & Education 65 (2013) 64–76

Contents lists available at SciVerse ScienceDirect

Computers & Education

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

Understanding the effects of professors' pedagogical development with Clicker Assessment and Feedback technologies and the impact on students' engagement and learning in higher education

Jae Hoon Han [a,*], Adam Finkelstein [b]

[a] Project Management Office (PMO), McGill University, 1450 – 688 Sherbrooke West, Montreal, Quebec, Canada H3A 3R1
[b] Teaching and Learning Services (TLS), McGill University, Montreal, Quebec, Canada H3A 0C8

ARTICLE INFO

Article history:
Received 26 February 2012
Received in revised form
27 January 2013
Accepted 1 February 2013

Keywords:
Improving classroom teaching
Interactive learning environments
Pedagogical issues
Post-secondary education
Teaching/learning strategies

ABSTRACT

Clicker Assessment and Feedback (CAF) is an instructional assessment and feedback strategy that is incorporated with interactive technologies, often referred to as clickers. Several thousand colleges and universities across Europe and North America have adopted CAF as a strategy in their classrooms. This study has three major objectives. The first objective is to discuss the development of an instrument used to assess and investigate students' perceptions of CAF tools. The second is to examine the effects of university professors' CAF development on student perceptions of CAF. The third is to investigate the impact of professors' CAF methods on student learning and engagement. In this study the CAF project was initiated to enhance students' engagement in undergraduate courses by supporting CAF development to university professors at a large, publically-funded University. Professors ($n = 74$) and students ($n = 5459$) volunteered to participate over this four-semester long project. Principal Component Analysis (PCA) was performed to explore students' perceptions of CAF efficacy. Multivariate Analysis of Variance (MANOVA) was used to explore the relationship between professors' CAF development, their use of CAF in formative or summative assessment and students' perceptions of CAF. The results demonstrate that 1) students perceive the use of CAF tools as having an impact on their engagement and learning, 2) increased CAF development by professors impact on students' perceptions of CAF, and 3) professors' use of CAF for formative assessment is more influential than summative assessment on students' perceptions of engagement and learning. This study suggests that CAF is most effective for student engagement and learning if it is supported by appropriate CAF development of professors and their subsequent formative use of CAF during teaching.

1. Introduction

Clicker Assessment and Feedback (CAF) is an easily adopted teaching approach for higher education (Bruff, 2009). CAF is also referred to as a Classroom Communication System (CCS), Student Response System (SRS), Audience Response Technology (ART), or clickers. CAF refers to inquiry-based teaching strategies coupled with a clicker technology system, a wireless system that enables professors to ask questions and have students respond using hand-held devices (clickers). The questions and the results summarizing student responses can be presented simultaneously on the classroom's projector (see Fig. 1).

Many studies have investigated the effects of CAF on various aspects of students' classroom experiences. The educational benefits of CAF can be summarized in terms of the aspects of teaching and learning affected: CAF enables professors' contingent teaching (McKeachie, 1990) by providing real-time feedback and assessment, as students anonymously participate in classroom activities (e.g., peer teaching). Professors can assess student understanding, and students can assess their own understanding of a concept (Kay & LeSage, 2009). Nobel Laureate Weiman (2010) summarized the use of CAF as having "… *a profound impact* on the educational experience of students" (p.186).

The *profound impact* on student learning is well documented in many CAF studies (e.g., Kay & LeSage, 2009). CAF can have a positive influence on students' emotional, motivational, and cognitive experiences in the classroom (e.g., Simpson & Oliver, 2007). Based in part on

* Corresponding author. Tel.: +1 514 398 4802 (Office); fax: +1 514 398 4718.
E-mail addresses: jaehoon.han@mcgill.ca (J.H. Han), adam.finkelstein@mcgill.ca (A. Finkelstein).

0360-1315/$ – see front matter
http://dx.doi.org/10.1016/j.compedu.2013.02.002

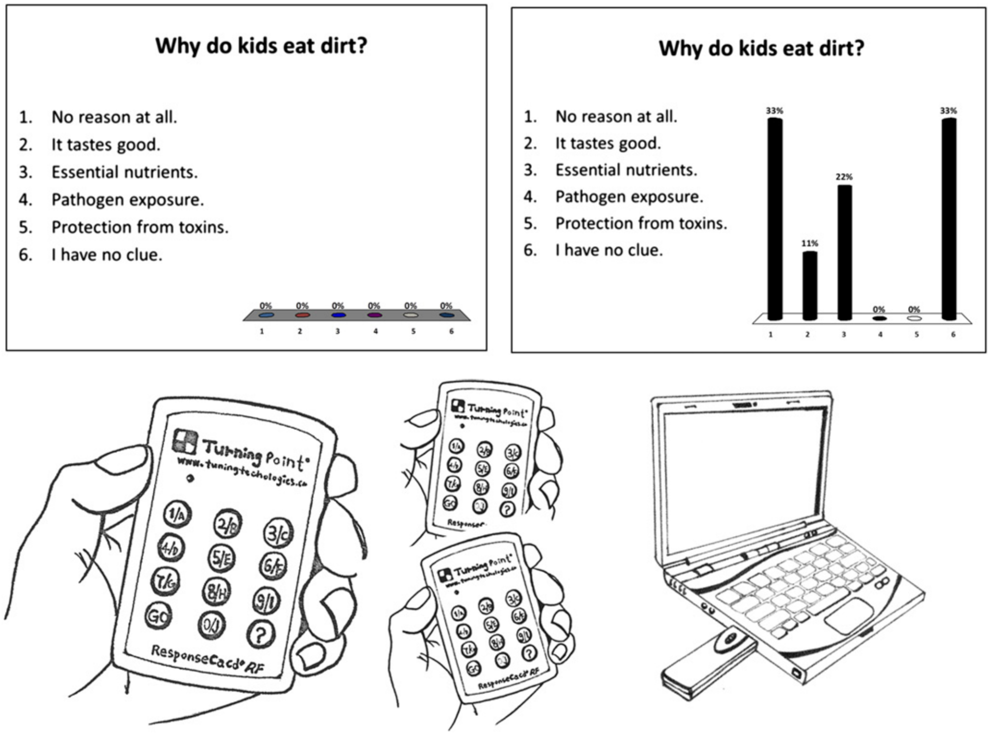

Fig. 1. Clicker Assessment and Feedback (CAF) mechanism. Professors can project questions on a wall or screen in their classrooms. Students can then respond to the questions on their hand held devices (clickers). A receiver is connected to a laptop – Receivers can gather students' responses and transform their responses into digital data. Professors can then display the results of the students answered if they choose.

the support of these findings, CAF has been used on over a half-million students in several thousand colleges and universities in Europe and North America (Steinberg, 2010, p. A14). However, two issues with CAF have been recognized. One is the validity and reliability of the instruments used to gather data on CAF use have been criticized, and thus, questions regarding the credibility of these studies have surfaced (e.g., Kay & LeSage, 2009). Most CAF studies (e.g., Bunce, VandernPlas, & Havanki, 2006; Hoekstra, 2008) measured students' perceptions of CAF as engagement and learning with only one or two items. This has led to concerns of reliability and validity (DeVellis, 2012). Another concern is that the relationships between professors' teaching and students' learning have not been adequately explored in CAF studies in higher education (Offerdahl & Tomanek, 2011). However, the literature on university teaching consistently demonstrates the educational effects of CAF on various aspects of student learning (e.g., Bruff, 2009; Simpson & Oliver, 2007). CAF studies have tended to focus on the effects of adopting CAF on students by comparing CAF and traditional lectures rather than exploring a holistic interaction among professor's teaching approaches, CAF, and student learning (Fies & Marshall, 2006; Kay & LeSage, 2009). The results of these studies do not provide the evidence of which components of CAF approaches are associated with which educational benefits and challenges (Fies & Marshall, 2006; Penuel, Boscardin, Masyn, & Crawford, 2007).

This study aims to respond to some of the criticism of CAF studies by exploring students' perceptions of CAF using Principal Component Analysis (PCA). PCA is a method of analysis that is a psychometrically sound technique with established credibility that avoids factor indeterminacy (Stevens, 2002). This study also intends to investigate the relationship among professors' pedagogical development using the technology, their use of technology, and students' perceptions of the impact of CAF on their learning, which is considered a *missing link* in the technology evaluation literature in higher education (Offerdahl & Tomanek, 2011; Simpson & Oliver, 2007).

2. Literature review

The initial clicker device prototype was designed and produced by Carpenter (1950a, p. 33, 1950b, p. 20) at Pennsylvania State University. Approximately 47 years later, Mazur (1997) wrote his seminal work, *Peer Instruction*, which resulted in the rapid adoption and implementation of CAF at several thousand universities and colleges in Europe and North America (Steinberg, 2010, p. A14). This may have been a result of the research of Mazur's Harvard Physics group (Crouch & Mazur, 2001; Fagen, Crouch, & Mazur, 2002; Larsy, Mazur, & Watkins 2008), which consistently highlighted the benefits of CAF on various aspects of students' experience. CAF was found to enhance student engagement with content and with peers in the classroom and impact learning by helping to address important misconceptions (Crouch & Mazur, 2001). Teachers of physics and virtually all other disciplines (Science, Technology, Engineering, Mathematics – STEM, Medicine, Social Sciences, and Arts and Humanities) began developing and implementing their own customized instructional strategies using CAF, such as class-wide discussion (Nicol & Boyle, 2003), technology-enhanced formative assessment (Beatty & Gerace, 2009), and the question-sequence method (Reay, Li, & Bao, 2008).

The growing body of CAF studies reveals various effects of CAF on student experiences in the classroom across virtually all disciplines in higher education: increasing attention (Latessa & Mouw, 2005), attendance (Moredich & Moore, 2007), interaction (Johnson & Meckelborg, 2008), teamwork (Stevenson, 2007) motivation (Van Dijk, Van Den Berg, & Van Keulen, 2001), positive emotion and participation (Stowell & Nelson, 2007), engagement and metacognition (Campbell & Mayer, 2009) and learning (Mayer et al., 2009). However, literature reviews (e.g., Kay & LeSage, 2009; Penuel et al., 2007) have strongly criticized most studies' disregard for the validity and reliability of the instruments used and have suggested that the basis for understanding the effects of CAF on students has been mostly *anecdotal* data.

In order to review the validity and reliability concern raised from previous literature reviews (Kay & LeSage, 2009; Penuel et al., 2007), 27 of 184 published journal CAF studies were selected in order to analyze how they used items (e.g., validity and reliability) when measuring learning and engagement with CAF in higher education. Reviews on the items were examined in the following themes: scales, reliability, validity, participant numbers, demographics, and the discipline of the instructions described in CAF studies. The CAF studies were found from searching academic databases (i.e., Scopus, PsyInfo, and ERIC) and the bibliography of CAF studies (Bruff, 2009) by using the keywords (e.g., Clicker, Audience Response Systems, Classroom Communication Systems). Detailed information regarding 27 studies reviewed is presented in the Appendix.

Upon reviewing the items used in 27 CAF studies, four noticeable themes emerged: 1) Most studies ($n = 24$) did not report or specify either the validity or reliability issues in items used; Only three studies (MacGeorge et al., 2008; Rice & Bunz, 2006; Siau, Sheng, & Nah, 2006) reported both the reliability and validity; 2) Most studies ($n = 23$) developed only one item to measure only one scale, which might raise a question of the item validity used; 3) Detailed descriptions of the items developed were not provided ($n = 23$); 4) Results of the most studies ($n = 21$) were reported on only one course *within* a discipline rather than multiple courses *across* disciplines. The five most frequently measured scales in those studies were Learning ($f = 14$), Understanding ($f = 12$), Enjoyment ($f = 10$), Involvement and Participation ($f = 8$), Attention ($f = 7$). Only five studies reported the process of the development of the instrument while others did not report or specify the development process. Demographics indicated that most participants were undergraduate and graduate students. Only three studies had different participants, medical residents (Latessa & Mouw, 2005; Miller, Ashar, & Getz, 2003; Pradhan, Sparano, & Ananth, 2005).

Among the three studies reporting the validity and reliability, two studies (Rice & Bunz, 2006; Siau et al., 2006) were conducted in specific courses with a relatively small number of participants (each 26 undergraduate and 46 graduate students) and within one specific discipline (i.e., Communication and Engineering Studies). MacGeorge et al. (2008) respond to a critique from these previous literature reviews (Kay & LeSage, 2009; Penuel et al., 2007) by providing relatively sound construct validity and reliability in the Audience Response Technology Questionnaire (ART-Q). It consists of 6 topics, 40 questions (ease of use, attendance, appraisal/learning, enjoyment, preparation/motivation, and negative grade). It was used with a total of 854 students across a number of disciplines (Liberal Arts, Consumer and Family Science, Technology, Agriculture, and Management). There are, however, still a number of concerns with the ART-Q. One concern of the ART-Q is that it was not fully validated across all disciplines, and therefore, may not be applicable or credible for assessing and comparing the effects of CAF across different university settings. A second concern is with its length (at 40 questions), which may make it difficult to implement across multiple courses and multiple institutions, especially during class time.

One of the issues mentioned in studies exploring CAF use in the classroom is that the relationship between professors' teaching factors and students' experiences had not been adequately studied (e.g., Offerdahl & Tomanek, 2011; Simpson & Oliver, 2007). The literature (e.g., Blood & Neel, 2008; Draper & Brown, 2004) suggests that CAF use provides opportunities for professors to re-examine and alter their teaching approaches as they incorporate CAF, rather than implementing a relatively new technology (CAF) without changing their teaching approaches. Many professors successful with CAF have subsequently developed their teaching knowledge and skills to integrate CAF into their teaching approaches to enhance students' classroom learning experiences (Simpson & Oliver, 2007). Draper and Brown (2004) clearly showed a positive correlation between professors' experience in CAF approaches and students' perceived satisfaction. These results and their implications are congruent with the findings of teaching development (intervention) studies in higher education (e.g., Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010; Weimer & Lenze, 1991). Teaching development has an impact on various outcomes (professors' thought processes and actions, student learning, and institutional change), and over the long-term has more influence on these outcomes than short-term interventions or no interventions at all (Weimer & Lenze, 1991). These studies strongly support the idea that teaching development that is focused on the use of Clicker Assessment and Feedback (CAF) could have an impact on professors' thought processes and actions as they incorporate CAF into their teaching (Offerdahl & Tomanek, 2011; Stes et al., 2010). This change in teaching approach may successively influence students' perceptions of learning and the quality of their learning outcomes (Trigwell, 2010).

Another under-represented issue in the CAF literature is the relationship between professors' specific use of CAF (i.e., for formative or summative assessment) and how students perceive its use in their learning (MacGeorge et al., 2008; Simpson & Oliver, 2007). Most studies examine the comparative effects of using CAF on various aspects of student learning compared to not using CAF in the teaching process (Campbell & Mayer, 2009), thereby claiming to demonstrate the educational benefits of the use of CAF (Kay & LeSage, 2009). Only a few studies have examined the relationship between professors' use of CAF for formative or summative assessment and students' experiences. For example, James (2006) and James, Barbieri, and Garcia (2008) found that there were no significant differences between students' learning when CAF was coupled with formative as opposed to summative assessment for introductory physics courses. Despite these results, an interesting finding was that student groups using CAF for formative purposes participated more and were more engaged in peer instruction activities than their counterparts using CAF for summative purposes (James, 2006). The literature has yet to examine the effects of professors' use of CAF for formative or summative assessment on students' perceptions of CAF's utility across disciplines. Moreover, no attention has been given to the relationships between professor characteristics, teaching factors (professors' teaching development and ways of using CAF) and students' perceptions of the utility of using CAF in the classroom.

3. Research questions

The following research questions emerged based on a literature review on CAF studies:

3.1. What is the nature of the underlying dimensions (factors) of students' perceptions of CAF utility in the classroom?

3.2. What are the relationships between professors' CAF development, their methods of using CAF (formative or summative), and students' perceptions of CAF utility in the classroom?

3.2.1. Are there significant differences in students' perceptions of CAF utility based on professors' CAF development experiences?

3.2.2. Are there significant differences in students' perceptions of CAF utility in relation to professors' formative or summative assessment methods of using CAF?

4. Methods

4.1. Context of the study

A large (re)design project, entitled *Enhancing Student Engagement Using Clicker Technology*, was initiated by a large publically-funded University. The two major goals of this project were to use clicker technology to: (1) to promote an active mode of learning by embedding inquiry-based practices and peer interaction into the classroom experience. (2) To provide real-time feedback and formative assessment to students to increase their engagement and learning. In order to achieve these goals, a primary focus was given to support university professors while they adopted a learner-centered teaching approach in conjunction with CAF use. The teaching and learning support unit designed the CAF development program for professors, which was delivered through two workshops (one of three and a half to four hours, the other significantly shorter). Before teaching a course using CAF, university professors were highly encouraged to attend the initial four-hour workshop, which focused more on pedagogical development than technical. Professors were required to bring their own laptops as well as questions into the first CAF development session.

The first part of the CAF workshop mainly focused on the pedagogical perspectives on using CAF. It began with a focus on principles of Course Design and Teaching (Saroyan & Amundsen, 2004) to help the professors understand the nature of course design in higher education followed by how to design effective CAF questions linked to the professor's learning goals and how to design and develop in-class activities to enhance student learning and engagement. Also, hands-on instruction was offered to present CAF examples (i.e., format, types, and effects), from existing courses in their discipline. The second part of the workshop was offered to provide a hands-on session, with the professors, for running CAF sessions. In this session, professors installed the application into their laptops and tested them out with each individual response set; while professors created their own CAF questions or converted their initial questions into CAF questions. Professor presented and discussed their CAF questions with other participants and educational developers; all participants provided their feedback to the CAF questions. Participants discussed the development of relevant answers for students to choose as part of the CAF activity (and the complexities of creating those options). A great deal of time was devoted to issues involved with using CAF for formative feedback as part of a learning activity. Lastly, the workshop also focused on how to help professors interpret student in-class responses and how to provide effective feedback.

The second workshop was shorter in length, occurred in the beginning part of the teaching term and was designed to help improve professors' understanding of using CAF in their courses. Professors were required to bring their laptop and data gathered from student responses in class. This workshop focused on how to organize students' CAF data and create reports for themselves and their students to become familiar with responses across multiple classes. Attendance at the first workshop was a prerequisite for attending the second workshop (Table 1).

Two types of questions were distributed to participants for understanding teaching and technology information about the professors before the workshops: 1) professors' overall years of teaching experience, 2) the section and level of the course offering, 3) the number of students in the courses, 4) experiences of using audio and video technology in teaching, and 5) any experiences of using CAF as a professor in the classroom. Of 89 participants, 74 indicated that they had no experience using CAF in their classrooms prior to the workshops. The average years of teaching experience was 5.76 ($SD = 8.28$) and the average number of students was 173. 12 ($SD = 178.21$). Most professors ($n = 72$) reported that they used one of the productivity software suites (e.g., MS-Office and iWork) as both visual and audio technology in their teaching; while 25 professors didn't use any audio technology in their classrooms. Professors participated only once in the research study during the project period (2008–2010). Professors' demographic information is shown in Table 2.

4.2. Instrument: Clicker Assessment and Feedback Questionnaire (CAF-Q)

The instrument, called Clicker Assessment and Feedback Questionnaire (CAF-Q), was designed by the teaching and learning support unit based on a literature review of items used in 27 CAF studies. A five point Likert-type scale was used for developing the instrument, which ranged from 1 (strongly disagree) to 5 (strongly agree). The goal of developing the CAF-Q was to be able to assess a large body of students' perceptions ($N = 6029$) of CAF using a sound instrument within and across the institution. A decision was made to gather the data in class using clickers in order to get the highest response rates possible – particular care was given to the length and focus of the instrument.

Table 1
CAF pedagogical development workshops for supporting professors teaching with CAF.

Workshop	Time/Duration	Learning goals	Required
1st workshop	Before the course (3.5–4 h)	■ Design questions linked to learning goals ■ Design answers to guide questions ■ Create clicker questions, and run a session using software ■ Using questions to support strategies ■ Provide feedback to students ■ Prepare to use clickers in your class	Voluntary participation
2nd workshop	During the course (1–2 h)	■ CAF data analytics: Collecting students' CAF data and matching them with students ID numbers and names ■ Understand the differences between formative and summative feedback ■ Generating CAF reports ■ Selecting and publishing CAF reports ■ Using CAF data collected from courses to enhance students' engagement and learning	Voluntary participation but 1st workshop required

Table 2
Background information of participants (74 professors and 5459 students).

Professors	Group	Professors (n)
Years of teaching	0–5 years	52
	6–10 years	11
	11–20 years	4
	21+ years	7
Number of students (n) in the class	0–49	7
	50–100	31
	101–500	27
	501+	9
Gender	Female	31
	Male	43

Professors & Students	Group	Professors (n)	Students (n)
Term(s)	Fall 2008	18	1426
	Winter 2009	21	1288
	Fall 2009	21	1799
	Winter 2010	14	946
Faculty	Agricultural and Environmental Sciences	11	484
	Arts/Humanities/Law	8	479
	Education	3	198
	Engineering/Architecture	5	177
	Medicine/Dentistry	4	111
	Management	20	1337
	Science	23	2672
Course level(s)	100 level	16	2397
	200 level	22	1550
	300+ level	36	1512

The conceptual framework of the items was based on Weston and Cranton's (1986) instructional strategy approach in higher education: instructional methods (i.e., feedback and assessment) and materials (i.e., clicker technology) should not be differentiated as two distinct concepts but rather should be used as an integrated instructional strategy. Specific technological perspectives (e.g., the ease of technology use as discussed by technology adoption theory), overall satisfaction, and recommendations were not considered when designing the CAF-Q. To ensure the content validity of the items (DeVellis, 2012), four content experts were involved in the item development process. One Learning Sciences professor had more than 30 years research and field experiences (e.g., development of the survey items, interview, and course teaching) of teaching and learning with technology in higher education; while the other three educational developers had 15–20 years of field and research experiences.

As previously noted (see Appendix), the scales used in the 27 CAF studies were reviewed by two educational developers: the most frequently used scales of other CAF studies are Learning ($f = 14$), Understanding ($f = 12$), Enjoyment ($f = 10$), Involvement and Participation ($f = 8$), Attention ($f = 7$), Attendance and Fun ($f = 6$), Interest ($f = 5$), Knowledge and Preparedness ($f = 4$), Stimulation ($f = 3$), Confidence, Clarification, and Critical Thinking ($f = 2$). Duplicated scales (e.g., Fun and Enjoyment, Involvement and Participation) were adjusted to one scale (e.g., Enjoyment and Involvement), although all scales from studies were considered into the measuring student learning and engagement when using CAF. To refine these items, based on Fredricks, Blumenfeld, and Paris's (2004) literature review on the relations between engagement and motivation, the scales and their related-items were recursively reviewed. Also, student cohorts participated in a pilot test a few times and they provided the internal committee with feedback about the items. In this regard, Attention, Attendance, and Involvement were inclusively categorized as the Engagement scale; while Assessing Learning, Understanding, Knowledge, Learning experience were categorized as Learning. These scales and items were reviewed and given feedback by two outside experts in North America (one educational developer in a large size university and one researcher in a medium size university whose expertise is with the topics of engagement and learning in higher education). In addition, the practicability of the survey was considered by the feedback from the participant professors – a large set of questions might bother their final class in the courses with the possibility of class time running over; hence the relatively smaller number of items were selected and decided upon the internal committee to use for conducting a survey within the institution.

Item reliability for two scales was examined for the questionnaire used in this study. As seen from Table 3, the Cronbach's α value for two scales is 0.80; according to George and Mallery (2012), the Cronbach's α value with 0.80 was considered as showing a good internal consistency of the items. Total statistical value of the items for Scale Mean, Correlation, Squared Multiple Correlation, Cronbach's α, and standardized Cronbach's α consistently showed a relatively good reliability of items (see Table 3). More specifically, Cronbach's α if the item deleted (see Table 3) was also used to examine the reliability of each item in the questionnaire, which ranged from 0.70 to 0.77. Compared to each α of two scales (each 0.8 in Engagement and Learning) they were still consistently reliable. This result indicated that Cronbach's α of two scales would be less reliable if an item was removed from either scale (the α would decrease 0.77–0.70). Therefore all seven items in the questionnaire were retained.

4.3. Data collection and instruments

Of 89 university professors who participated in the CAF development sessions from 2008 to 2010 academic years over 4 semesters, 83.16% professors ($n = 74$) volunteered to participate in this study. Of 6029 students registered in those professors' courses, 90.55% students

Table 3
Summary of total statistics for the questionnaire.

Item	*M*	*SD*	σ^{2a}	R^2	α^a	α	Standardized α
Engagement[b]						0.80	0.80
E1. Having to respond to clicker questions and feedback improved my attention.	3.98	0.92	3.06	0.39	0.74		
E2. The use of clicker feedback encouraged me to be attentive to classes.	3.90	0.94	2.99	0.39	0.72		
E3.When using clicker feedback used in a class I felt more involved.	3.66	1.03	2.67	0.42	0.70		
Learning[b]						0.80	0.80
L1. The use of clicker feedback improved my understanding of the course content.	3.87	0.97	6.41	0.37	0.75		
L2. Clicker feedback helped me focus on what is should be learning in the course.	3.81	0.98	6.26	0.40	0.74		
L3. I found using clicker feedback in the class worthwhile to my learning experience.	3.49	1.05	6.23	0.33	0.77		
L4. Viewing the class answers and feedback to clicker questions helped me assess my understanding of the course content.	3.76	1.07	5.88	0.41	0.73		

[a] If the item deleted.
[b] Factor.

(n = 5459) agreed to participate in the study. Participants' background information is shown in Table 2. Data were collected over four semesters (Fall 2008, Winter 2009, Fall 2009, and Winter, 2010). To collect student data for courses in each semester, four educational developers visited classrooms at the end of the last class (13th, or 14th course week depending on the course schedule). Before students' participation, educational developers presented the purpose of the survey and how to respond the questions using the clickers. Before proceeding, one testing question was presented to check the functionality of each student's individual clicker set. The entire questionnaire, of seven items, was administered one question at a time using clicker software in the classroom. Each student individually responded using their clicker. To keep the results anonymous, all identification data gathered by the system (e.g., barcode ID of the clicker) was made anonymous. In addition, educational developers did not show the results of the question answers in class, or to the instructor. All data were anonymously saved in a digital file. Teaching related data (e.g., course offering name and level of the course) were collected from university' course registration and information system with the professor participants' permission of use.

Information about professors' CAF development and use of CAF were collected from the CAF development documents, course information systems, and phone or face-to-face interviews conducted by four educational developers. Demographic information is shown in Table 2. SPSS 20.0 was used to analyze the data in this study.

4.4. Data analysis and variables

Using data that had been collected over four semesters, Principal Component Analysis (PCA) was employed to investigate the underlying factors of the questionnaire collected within and across the institution. The major goals of PCA are to examine the pattern correlations among the variables and reduce those variables into a small set of factors (Tabachnick & Fidell, 2012), which might provide useful information for further analysis of the factors PCA with other variables (Mulaik, 2010). In addition, using PCA could provide more determinacy than Common Factor Analysis (CFA) and can provide more useful information about the major features of data. All of which were appropriate for the goals of this study. We intended to examine and summarize the majority of the data and extract it into a small set of variables to further investigate the effects of professors' CAF development and use of CAF on students by using MANOVAs. However, it should be mentioned that each Factoring Analysis (FA) method has its own advantages and challenges regarding the purpose and data type related to the analysis employed in the study (Fava & Velicer, 1992; Mulaik, 2010; Tabachnick & Fidell, 2012) and that the debate is still ongoing for which factoring method would be the better fit for this specific data (e.g., Fava & Velicer, 1992; Velicer & Jackson, 1990). Using PCA could provide more determinacy than Common Factor Analysis (CFA), while Image Analysis (IA) could provide more generalizability and accuracy of the results than PCA and CFA (Mulaik, 2010). However, IA provides more considerable accuracy than any other FA methods and it is most appropriate for studies examining and generalizing universal and observable concepts rather than latent and unobservable variables (Mulaik, 2010). Considering that this study is to examine the latent, unobservable, and still developing variables, such as Engagement and Motivation, PCA method would be appropriate in providing relatively valid result, more so than CFA, but less than IA. Also, it should be noted that the differences within the factoring methods might be too small when analyzing large sample data (Tabachnick & Fidell, 2012).

A MANOVA was used to examine the effects of two independent variables (i.e., professors' CAF development and professors' use of CAF) and two dependent variables (i.e., student engagement and learning), which were obtained by PCA. As Tabachnick and Fidell (2012) indicated that the components extracted from PCA are orthogonal so using them as dependent variables in MANOVA could facilitate the process of interpretation of the results. In this study, a MANOVA was used to investigate not only the effect of each professors CAF development and use of CAF group differences on engagement and learning but also the interaction effect of two groups on student engagement and learning simultaneously. In addition, using a MANOVA can prevent Type I error inflation when using a multivariate test. Given that the main effects of the group found, post-hoc comparisons test was conducted to investigate the dependent variables (i.e., student engagement and learning in this study) were responsible for the results as well as the groups (i.e., CAF development and way of using CAF) were significantly different and on student engagement and learning.

Professors' CAF development was categorized into three levels (see Table 4): 1) *2 training sessions,* indicating that professors participated in both of the workshops offered, 2) *1 training session,* indicating that professors participated in only the first workshop offered, and 3) *No training,* indicating that, for that class, professors did not participate in any CAF workshops; these professors applied CAF approaches without any training. The other independent variable was professors' use of CAF, which was categorized into two types: Formative Feedback (FF) and Summative Feedback (SF). In order to categorize professors' use of CAF either as FF or SS, FF was operationalized as questions in class in which professors used the CAF approach for informal feedback *without* linking students' responses to their grades, whereas SF refers to classes in which professors linked CAF to grades (e.g., participation scores or points based on student CAF

Table 4
Descriptions of independent variables used in this study.

Independent variable	Professors (*n*)	Students (*n*)	Group	Description
Professors' CAF development	27	2174	2 training sessions	Professors participated in both of the two workshops offered before and during their use of CAF in courses.
	30	1699	1 training session	Professors participated in only the first workshop offered before using CAF in their courses.
	17	1586	No session	Professor did not participate in any CAF workshops before and during their use of CAF in courses.
Professors' use of CAF	47	2923	Formative Feedback (FF)	Professors and/or students reported that CAF was used by providing informal feedback *without* associating students' responses to their grades.
	27	2536	Summative Feedback (SF)	Professors and/or students reported that CAF was *somewhat* and *always* used by linking students' responses to their grades, such as attendance/participation bonus for students' final grades.

answers to questions). Of 74, 47 professors were categorized as using CAF as Formative feedback (FF), while remaining 27 were categorized as using CAF as Summative Feedback (see Table 4). The categories were set as mutually exclusive. Professors were categorized as being either FF or SF.

The η^2-index was used to calculate the effect sizes for both the independent and dependent variables. Based on Cohen's (1988, pp. 283–288) guidance – the table of convertor η^2 as a function of *F* test, $\eta^2 = 0.01$ ($F = 0.10$) was deemed a small effect, $\eta^2 = 0.06$ ($F = 0.25$) a medium effect, and $\eta^2 = 0.14$ ($F = 0.40$) a large effect. Bonferroni's adjustment alpha level of 0.025 was used to prevent the Type I error in performing the post-hoc comparison tests. Finally, an alpha value of 0.01 was used for all statistical analyses.

5. Results

5.1. Factor analysis

The factorability of the seven items was examined to determine whether further factor analysis was appropriate. First, the correlation matrix for the seven items was checked and revealed correlations between all items of at least 0.03. Second, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was conducted, revealing a value of 0.77, indicating a near "meritorious" level (Kaiser & Rice, 1974, p.114). Bartlett's test of sphericity was significant: χ^2 (21, $N = 5459$) = 11569.23, $p < .001$. The anti-image correlation matrix revealed that all items were over 0.5 and the communalities showed all items exceeding 0.5 (see Table 3). All tests indicated that the data were suitable for further factor analysis.

Principal Component Analysis (PCA) was used to explore the underlying factors for the seven items. The results revealed two factors with eigenvalues exceeding 1. Cattell's (1966) test was used to examine the appropriate number of factors and determined two. Parallel Analysis (Horn, 1965) was also conducted to investigate the number of factors using seven items with 5459 respondents, the results validated the two factors. Two factor solutions were also retained by Oblimin rotation and explained 64.33% of the total variance. Factors one and two explain 38.93% and 25.40%, respectively.

Overall, PCA revealed two factors. First, the factor measuring student learning included the following items: understanding content, assessing understanding, focusing on content, and usefulness of CAF for students' learning. Second, the student engagement factor included: increasing attention and being attentive, and involvement in the class. Tables 4 and 5 illustrate total statistics of the questionnaire: mean, mean if the item deleted, standard deviation, correlation, and Cronbach's alpha value for each item, and the coefficients and communalities for the two factors, along with their internal consistent reliabilities.

Table 5
The matrix of pattern and structure coefficients and communalities, correlations, and factor loadings for the questionnaire.

Factor	Item	Pattern coefficient		Structure coefficients		h^2
		1	2	1	2	
Engagement	E3.	**0.85**	0.02	**0.85**	0.19	0.73
	E2.	**0.84**	−0.02	**0.84**	0.15	0.71
	E1.	**0.83**	0.00	**0.83**	0.17	0.70
Learning	L3.	0.05	**0.80**	0.21	**0.81**	0.67
	L2.	−0.04	**0.79**	0.11	**0.78**	0.61
	L1.	−0.02	**0.75**	0.13	**0.75**	0.56
	L4.	0.02	**0.74**	0.17	**0.75**	0.56
Factor correlations						
Engagement				–		
Learning				0.2		–

Factor	Eigenvalue	% of variance	Cumulative %
1	2.73	38.93	38.93
2	1.78	25.40	64.33

Note: Substantial loadings are indicated as bold.

Table 6
Descriptive results: Student engagement and learning in relation to professors' CAF development and ways of using CAF technology.

Dependent variables	Ways of using CAF technology	*M* (*SD*)			
		No session	1 Session	2 Sessions	Total
Engagement	Formative	3.52 (0.52)	3.66 (0.70)	4.51 (0.64)	4.25 (0.76)
	Summative	3.24 (0.53)	3.55 (0.59)	4.40 (0.67)	3.50 (0.68)
	Total	3.25 (0.53)	3.60 (0.63)	4.48 (0.65)	3.85 (0.81)
Learning	Formative	3.44 (0.63)	3.86 (0.76)	4.00 (0.68)	3.93 (0.71)
	Summative	3.42 (0.86)	3.69 (0.80)	3.78 (0.81)	3.56 (0.84)
	Total	3.42 (0.85)	3.76 (0.79)	3.94 (0.71)	3.73 (0.80)

5.2. MANOVA

Based on the identification of the two factors, a 2 × 3 MANOVA was used to investigate the difference between professors' CAF development experiences, the types of feedback given to students and students' perceptions of their engagement and learning. In order to investigate the effects of each dependent variable (i.e., CAF development, CAF use of professors) on student perception of learning and engagement, post-hoc comparison analysis was conducted in this study. Descriptive and multivariate analyses results are shown in Tables 6 and 7.

Assumptions for multivariate tests were met for linearity (Bartlett's Test of Sphericity, $p < .001$) and multicollinearity (all VIFs <1.76). However, Box's *M* test indicated that the assumption of equalities of covariances was violated ($p < .001$); thus, Pillai's trace was used to examine the multivariate tests in this study. Multivariate tests indicated that there were significant effects of professors' feedback types and training on students' perceptions of the use of CAF in terms of their engagement and learning, Pillai's trace = 0.007, $F(2,5452) = 20.569$, $p < .001$, $\eta^2 = 0.01$, showing a small effect; Pillai's Trace = 0.233, $F(4, 10{,}906) = 359.129$, $p < .001$, $\eta^2 = 0.12$, showing a medium effect. Interaction effects on students' perceptions were not significant: Pillai's trace = 0.001, F (4, 10,906) = 1.865, $p > .114$, $\eta^2 = 0.00$.

Post-hoc test was conducted to examine which variables in this study (i.e., professors' CAF development and ways of using CAF) have a significant effect on student perceptions of engagement and learning. The analysis results of the professors' CAF development revealed a significant effect on students' perceptions of engagement: F (2, 5453) = 793.26, $p < .001$, $\eta^2 = 0.23$; and learning, $F(2, 5453) = 34.477$, $p < .001$, $\eta^2 = 0.01$. Pairwise comparisons revealed significant differences between students' perceived engagement between training groups (all $p < .001$), wherein the student groups whose professors had participated in two training sessions perceived higher levels of engagement ($M = 4.45$, $SD = 0.02$) versus those whose professors had participated in only one session ($M = 3.60$, $SD = 0.02$) and those whose professors had not participated in any training ($M = 3.38$, $SD = 0.04$). The 1-training-session groups also had higher levels than the no-training student groups. Pairwise comparisons also found significant differences between students' perceived learning for the different training groups (all $p < .001$), wherein the 2-training-sessions groups perceived higher levels of learning ($M = 3.88$, $SD = 0.02$) than 1-training-session groups ($M = 3.78$, $SD = 0.02$) and the no-training-session groups ($M = 3.43$, $SD = 0.05$).

Post-hoc analysis of the professors' use of CAF found significant differences in students' perceptions of engagement, $F(1, 5453) = 32.507$, $p < .001$, $\eta^2 = 0.01$, and of learning, $F(1, 5453) = 10.809$, $p < .001$, $\eta^2 = 0.00$. In other words, Formative Feedback (FF) groups reported higher levels of perceived engagement ($M = 3.90$, $SD = 0.03$) and learning ($M = 3.76$, $SD = 0.03$) as compared to Summative Feedback (SF) groups' engagement ($M = 3.72$, $SD = 0.01$) and learning ($M = 3.63$, $SD = 0.02$).

6. Discussion

One of the goals of this study was to explore underlying factors affecting students' perceptions of the utility of CAF and to develop a relatively credible instrument to measure such perceptions. The PCA analysis clearly found that students perceived the utility of CAF as having two discrete dimensions: engagement and learning. The learning dimension identified in this study was consistent with MacGeorge et al.'s (2008) learning/appraisal dimension. However, the engagement dimension differed from MacGeorge et al.'s (2008) in that engagement inclusively captured two dimensions (attention and motivation) as one.

6.1. The relationship between engagement and motivation

The relationship between engagement and motivation was an interesting result of this study, as the engagement dimension included some aspects of the motivational constructs described in MacGeorge et al. (2008). Other studies (e.g., Oliver, Tucker, Gupta, & Yeo, 2008; Reeve, Jang, Carrell, Jeon, & Barch, 2004) provided insight into understanding, both conceptually and practically, the relationship between engagement and various motivational constructs. Conceptually, engagement and motivation might not be seen as distinct concepts in many

Table 7
Multivariate and univariate analyses of variance for engagement and learning measures.

Source	Multivariate				Univariate							
					Engagement				Learning			
	df	*F*	*p*	η^2	*df*	*F*	*p*	η^2	*df*	*F*	*p*	η^2
CAF Training (T)	4	359.13	<.001	0.12	2	793.26	<.001	0.23	2	32.51	<.001	0.01
Ways of using CAF (W)	2	20.57	<.001	0.01	1	32.51	<.001	0.01	1	10.81	<.001	0.01
T × W	4	1.87	.11	0.00	2	2.16	0.12	0.00	2	1.36	.26	0.00

Note. Multivariate *F* ratios were generated from Pillai's statistic.

fields. Therefore, taking into account evidence from Fredricks et al. (2004), this study used one scale to assess engagement in order to increase construct validity. PCA analysis in this study suggested that students' perceptions of some aspects of motivational constructs could be captured into one *engagement* dimension, retaining students' perceptions of learning as a separate dimension as suggested by MacGeorge et al. (2008).

6.2. Impact of university professors' development on student learning and engagement

The second goal of this study was to investigate the relationship between university professors' CAF development and their students' perceptions of the utility of CAF. The results of this study provide compelling evidence that professors' CAF development has an impact on student perceptions of their engagement and learning when using CAF: the more the professors participated in CAF development the more CAF teaching experience they had, and the more students perceived that the use of CAF increased their engagement and learning. Stes et al.'s (2010) and Weimer and Lenze's (1991) reviews provide significant insight for understanding the results of this study, with the finding that the longer professors engaged in university teacher training (intervention or development), the more influence they had on student learning. Consequently, more time spent training (2 CAF development sessions in this study) had a more positive effect than less time (1 CAF development session in this study) on student learning (Stes et al., 2010). This study thus provides empirical support for Stes et al. (2010) and Weimer and Lenze's (1991) hypothesis: professors' CAF development showed a medium effect on both student engagement and learning ($\eta^2 = 0.12$), and a large effect on student engagement ($\eta^2 = 0.23$). Post-hoc comparisons also found that the effect of 2 sessions of CAF development on student engagement and learning was significantly higher than that of 1 session of CAF development, while 1 session of CAF development also had a significantly stronger effect than no training (all $p < .001$). Thus, these results suggest that longer-term CAF teacher training leads to greater impacts on student perceptions of the utility of CAF for improving their engagement and learning.

It should be noted that this study did not take into account previous pedagogical knowledge of the professors' use of CAF. Professors that attended the CAF development sessions may have already developed the pedagogical knowledge and may be aware of how to effectively integrate CAF in their teaching. However, if that were a pervasive concern, then their feedback on the development process would have been negative – quite the contrary, their feedback on these sessions were extremely positive. A majority of the professors that were part of the study were in the beginning of their careers (0–5 years' experience; see Table 2) making their background pedagogical knowledge less likely to be an impacting factor – their motivation and interest may have had an impact on their successful use of CAF.

6.3. Relationship between a way of using CAF on student learning and engagement

The final goal of this study was to explore the relationship between professors' ways of using CAF (for formative or summative assessment) and student perceptions of CAF utility. Post-hoc tests clearly revealed that students in classes using CAF for FF (formative feedback) rated their levels of engagement and learning significantly higher than did students in classes using CAF for SF (summative feedback) (all $p < .001$). This study partly confirms James' (2006) and James et al.'s (2008) findings that professors' formative use of CAF has more impact than professors' summative use of CAF on student perceptions of engagement. Concurrently, this study adds a unique finding to a growing body of CAF literature: professors' formative use of CAF also has more influence than their summative use of CAF on students' perceptions of their learning in the classroom. Studies on assessment and feedback in higher education (e.g., Knight & Yorke, 2003; Weston, Le Maistre, McAlpine, & Bordonaro, 1997) consistently suggest that formative feedback strategies outperform the counterpart on various student experiences or outcomes. Therefore, it could be argued that formative assessment and feedback strategies might have more effect than summative counterparts on students within specific settings (i.e., classrooms using the CAF approach in this study).

7. Limitations and recommendations for future research

Although this study is an important first step towards a comprehensive understanding of students' perceptions of CAF, professors' CAF development, and how professors use CAF, our work has a few limitations. This study only investigated students' perceptions of their engagement and learning, which may or may not be related to the learning approach and the quality of student learning outcomes (Biggs, 2003; Entwistle, 2010) when using CAF in classrooms. In addition, students' perceptions of CAF might not be directly affected by the professors' methods of using CAF, since the professors' characteristics might have more impact than any other factor (e.g., the use of CAF in this study) on students' ratings of their classroom learning and experiences (d'Apollonia & Abrami, 1997). Therefore, the future studies might investigate the relations among professors' teaching approaches, students learning approaches, and with their use of CAF by employing a longitudinal classroom observation to examine the effects of CAF on students. In addition, it should be noted that relatively small numbers of items were used to examine student engagement and learning, which might threaten the construct validity and reliability for measurement. Finally, it should also be noted that the effects of professors' CAF development and use of CAF might have *a short term* and *limited* impact on student engagement and learning (i.e., Novelty effect; Clark, 1983) since CAF approach was only introduced as a new technology at the institution in 2008. Future studies could examine *a longitudinal* impact of professors' CAF development and using CAF on student engagement and learning.

8. Conclusions and implications

This study has highlighted the critically important relationship between students' perceptions of CAF utility, university professors' CAF development, and professors' use of CAF. Our work has led us to conclude that when using CAF in the classroom, 1) students perceive CAF as useful for two discrete dimensions: engagement and learning, 2) professors' longer-term CAF development has more impact on student perceptions of CAF's utility for their engagement and learning, and 3) formative feedback with CAF has more influence than summative feedback with CAF on students' perceptions of CAF utility in terms of its impact on their engagement and learning.

This study contributes a more nuanced understanding of the consequences of the use of CAF in the field of technology and assessment in higher education. First, this study clearly demonstrates the significant effects of professors' CAF development on students' perceptions of

CAF. This is supported by other CAF literatures (e.g., Bruff, 2009; Wieman, 2010) that suggest that pedagogical development activities might play a crucial role of adopting CAF in the classroom. Despite this, there were no significant discussions or studies about how to design and implement a faculty development program when using a new emerging technology (i.e., CAF) and what the role of a development program is, in teaching and learning. The evidence from this study suggests that university or college teaching development staff (educational developers) should appropriately and proactively design and disseminate new assessment and feedback teaching development strategies to enhance professors' adoption of new emerging technologies for their teaching practices (Knight & Yorke, 2003). Especially, as other faculty development studies have shown, a professors' training would have more impacts on students when using CAF in the classrooms. Readers are encouraged to contact the authors regarding the CAF development program that was implemented in an effort to share best practices and further refine the development program with other institutions.

Second, this study provides the CAF-Q as an instrument that can be used to assess students' perceptions of the efficacy of CAF technology. As the CAF-Q was validated across all undergraduate class-levels and disciplines, it is thus a relatively credible instrument for use in the context of any higher education institution (colleges or comprehensive universities). Moreover, the CAF-Q demonstrates a very high factorability, revealing high communalities, pattern/structure co-efficient, and internal consistency, which might be caused by using relatively small numbers of items. However, the CAF-Q could be appropriate for practitioners to capture students' perceptions for the efficacy of CAF in relation to other factors (e.g., professor and student background characteristics).

Finally, this study provides empirical evidence that professors' use of CAF for formative assessment is more influential on student engagement and learning than its use for summative assessment. However, as seen in the results, some university professors still use CAF for implementing summative assessment strategies, even though they have been advised in the sessions of the benefits of using it for formative assessment and feedback. There is thus a need for future studies on the reason why some professors are still resistant to providing more opportunities for formative feedback, perhaps using single or multiple case studies (Yin, 2009). Additionally, studies examining the relationship between professors' thoughts about teaching and technology and their teaching behavior when using an emerging approach would provide insight into their reluctance to avoid formative assessment when using CAF. These questions should be assessed using multiple validated methods (Offerdahl & Tomanek, 2011) in order to establish a critical base of CAF studies that are systematic, detailed, and rigorous (Kay & LeSage, 2009).

Given the nature of complex relations among professors' pedagogical development of CAF, the ways of using CAF, and student learning and engagement, this study provides a snapshot for understanding the effects of professors' development with CAF as a new technology, and their ways of using CAF on student learning and engagement. Most importantly, this study empirically demonstrates that pedagogical development matters when a new technology for teaching and learning is adopted for the higher education classroom.

Acknowledgments

This study was financially supported by the Arthur and Dawn Weston Fellowship, McGill International Doctoral Awards (MIDAs), Teaching and Learning Services (TLS), McGill University, Montreal, Canada. The authors acknowledge Professor Cynthia Weston, the director of TLS, for her great feedback on this manuscript and Pierre-Andre Vungoc and Effi Kaoukis for their support with the data collection. The authors also thank the reviewers for their constructive comments and suggestions.

Appendix

Results of review on 27 CAF studies regarding the scale, item, participants, course level, discipline, and reliability and validity.

Authors	Scales (Number of items used to measure scale)	Item development description	Items/Scales	Participants (n)	Course level	Discipline (or Course)	Reliability or Validity
Bojinova and Oigara (2011)	Ease of Use (1), Interesting (1), Engagement (1), Understanding (2), Involvement (1), Anonymity (1), Participation (1), Feedback (1), Misconceptions (1), Discussion (1), Beneficence (1), Recommendation (1), Enjoyment (1), Satisfaction (1), Grade (1)	Not specified or reported	16 items/Likert scale ranged from 1 to 5	40	Undergrad.	Principles of Microeconomics, Physical Geography	Not specified or reported
Boyle and Nicol (2003)	Conceptual Understanding (6) – Understanding (2), Involvement (2), Remember (1), Preference (1) Interaction and Discussion (6) – Attention (1), Checking Information (5) Motivation (5) – Attention (1), Confidence (1), Enjoyment (1), Preference (1), Helpful to Teacher (1)	Conducted focus group interviews with reorganizing the core issues based on the results of the interviews	6 items/Likert scale ranged from 1 to 5	117	Not specified or reported	Engineering Mechanics	Not specified or reported
Brewer (2004)	Thinking (1), Knowledge (1)	Not specified or reported	2 items/Likert scale ranged from 1 to 5	270	Undergrad.	Introductory Biology	Not specified or reported
Bunce et al. (2006)	Enjoyment (1), Learning (1), Preparedness for Quizzes (1), Preparedness for Exams (2), Confidence (1)	Not specified or reported	5 items/Likert scale ranged from 1 to 5	41	Undergrad.	General, Organic, and Biochemistry	Not specified or reported
Cardoso (2011)	Motivation/Interest (1), Participation/Involvement (1), Self-Assessment (1), Feedback (1), Interaction (1), Learning (1), Grades (1), Recommendation (1)	Not specified or reported	8 items/Likert scale ranged from 1 to 5	30	Second Language Learners	English as a Second Language (ESL)	Not specified or reported
Corcos and Monty (2008)	Enjoyment (1), Learning (1), Clarity (1), Benefit (1), Satisfaction (1)	Not specified or reported	5 items/Likert scale ranged from 1 to 5	254	Undergrad.	Library Search Session	Not specified or reported
Crossgrove and Curran (2008)	Involvement (1), Attention (1), Understanding (1), Benefit (2), Affordability (1), Recommendation (1), Interaction (1), Motivation (1), Ease of Use (1), Participation (1)	Not specified or reported	10 items/Likert scale ranged from 1 to 11	229	Undergrad.	200 level Genetics (1 class)	Not specified or reported
D'Arcy, Eastburn, and Mullally (2007)	Enjoyment (1), Learning (1), Attendance (1), Preferences (1), Participation (1)	Not specified or reported	5 items/Likert scale ranged from 1 to 3	150	Undergrad.	200 level Plant Pathology	Not specified or reported
DeBourgh (2008)	Testing Knowledge (2), Understanding (4), Benefit (1)	Not specified or reported	7 items/Likert scale ranged from 1 to 5	65	Undergrad.	Nursing Therapeutics	Not specified or reported

(continued on next page)

(*continued*)

Authors	Scales (Number of items used to measure scale)	Item development description	Items/Scales	Participants (*n*)	Course level	Discipline (or Course)	Reliability or Validity
Fitch (2004)	Enjoyment (1), Satisfaction (1), Interest (1), Learning (1), Benefit (1), Attention (1), Involvement (1), General Preferences (3)	Not specified or reported	10 items/Likert scale ranged from 1 to 5	55	Undergrad. (200- and 300 -levels)	Communication Disorders	Not specified or reported
Gauci, Dantas, Williams, and Kemm (2009)	Engagement/Interest (1), Stimulating (1), Understanding (1), Learning, Involvement (1), Academic Skills (1), Attendance (1), Satisfaction (1), Preferences of Using CAF (4)	Not specified or reported	12 items/Likert scale ranged from 1 to 5	145	Undergrad. (200 level)	Physiology	Not specified or reported
Graham, Tripp, Seawright, and Joeckel (2007)	Participation (2), Grades (1), Self-Assessment (1), Peer Assessment (1), Formative Assessment (1), Understanding (1), Pacing (1), Mutual Awareness (1), Discussions (1), Interaction (1), Experiments (1), Exploration (1)	Literature/Related information reviewed	14 items/Likert scale ranged from 1 to 5	688	Undergrad.	Biology, Chemistry, Education, Marriage Family and Human Development, Physics, Psychology, Statistics	Not specified or reported
Greer and Heaney (2004)	Understanding (1), Class Preparedness (1), Retention (1), Attendance (1), Fun (1), Interaction (1), Efficiency (1), Overall Satisfaction (4), Participation (1), Recommendations (2), Benefit (1), Enjoyment (1)	Not specified or reported	Total 16 items: 12 items/Likert scale ranged from 1 to 5 (4 items/Yes or No)	582	Undergrad.	Geoscience	Not specified or reported
King and Robinson (2009).	Usefulness, Overall Benefits (other scales with the number of the item used are not specified or reported)	Not specified or reported	13 items/Likert scale ranged from 1 to 5	250	Undergrad.	Engineering Mathematics	Not specified or reported
Latessa and Mouw (2005)	Fun (1), Attention (1), Learning (1), Satisfaction (1)	Not specified or reported	4 items/Likert scale ranged from 1 to 4	46	Faculty	Family Medicine	Not specified or reported
MacGeorge et al. (2008)	Ease of Use (3), Attendance (3), Appraisal/Learning (9), Enjoyment (9), Preparation/Motivation (4), Negative Grade (3)	Literature reviewed	31 items/Likert scale ranged from 1 to 5	854	Undergrad.	100-level: Introduction to Communication Theory, Forestry and Natural Resources 200-level: Organizational Leadership and Supervision	Ease of use (0.74), Attendance (0.86), Appraisal/Learning (0.84), Enjoyment (0.85), Preparation/Motivation (0.78), Negative Grade (0.86)
Miller et al. (2003)	Quality of the Session with CAF (2), Attention/Interest (1), Clarification (1), Benefit (1)	Revised thrice with using a focus group survey	23 items/Likert scale ranged from 1 to 5	283	Residents (Medicine)	Medicine	Not specified or reported
Mollborn and Hoekstra (2010).	Preference (1), Attendance (1)	Not specified or reported	2 items/Likert scale ranged from 1 to 5	350	Undergrad.	Introductory Sociology (i.e., Gender, Drug)	Not specified or reported
Morling, McAuliffe, Cohen, and DiLorenzo (2008)	Attention (1), Attendance (2), Enjoyment (1), Preparedness (1)	Not specified or reported	6 items/Likert scale ranged from 1 to 5	1290	Undergrad.	Introduction to Psychology	Not specified or reported
Poirier and Feldman (2007)	Preparedness (1), Applying (1), Fun (1), Interaction (1), Learning (1), Effectiveness (1)	Not specified or reported	6 items/Likert scale ranged from 1 to 4	865	Undergrad.	Introduction to Psychology	Not specified or reported
Pradhan et al. (2005)	Understanding (1), Ease of Use (1), Learning (1)	Not specified or reported	3 items/Likert scale ranged from 1 to 5	17	Residents (Medicine)	Medicine	Not specified or reported
Rice and Bunz (2006)	Training (5), Ease of Use (5), Perceived validity of responses (5), Fun (5)	Literature reviewed	20 items/Likert scale ranged from 1 to 5	46	Graduate	Research Methods, and Mediated Communication in Organizations	Overall (0.88/0.89), Training (0.91/0.92), Ease of Use (0.73/0.75), Validity (0.75/0.80), Fun (0.87/0.85)/ Conceptual Validity (comparing with other scales)
Roselli & Brophy (2006)	Time (1), Efficiency (1), Stimulating (1), Attention (1), Anonymity (1)	Not specified or reported	5 items/Likert scale ranged from 1 to 5	94	Undergrad.	Engineering	Not specified or reported
Shaffer and Collura (2009)	Involvement (1), Fun (1), Stimulating (1), Effectiveness (1), Understanding (2)	Not specified or reported	6 items/Likert scale ranged from 1 to 5	77	Undergrad.	Introduction to Psychology	Not specified or reported
Siau et al. (2006)	Interactivity (20, c.f., each 10 for individual and group) included Involvement, Engagement, Participation, Feedback, Self-Assessment (c.f., each 2 for individual and group), Perceived Ease of Use (3), Perceived Usefulness (3)	Literature reviewed for developing interactivity, the Technology Acceptance Model (TAM)	26 items/Likert scale ranged from 1 to 9	26	Undergrad.	Systems Analysis and Design	Cronbach's alpha for Interactivity (Individual-level: 0.86, group-level: 0.90), Perceived Ease of Use (0.96) and Usefulness (0.73).
Stein, Challman, and Brueckner (2006)	Learning (1), Understanding (1), Benefit (1), Clarification (1), Satisfaction (1)	Not specified or reported	5 items/Likert scale ranged from 1 to 4	76	Undergrad.	Anatomy and Physiology for Nursing	Not specified or reported
Williams and Boyle (2008)	Benefit (1), Learning (1), Understanding (1)	Not specified or reported	3 items/Likert scale ranged from 1 to 5	53	Undergrad.	Medicine	Not specified or reported

References

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*(11), 1198–1208. http://dx.doi.org/10.1037/0003-066x.52.11.1198.

Beatty, I., & Gerace, W. (2009). Technology-enhanced formative assessment: a research-based pedagogy for teaching science with classroom response technology. *Journal of Science Education and Technology, 18*(2), 146–162. http://dx.doi.org/10.1007/s10956-008-9140-4.

Biggs, J. B. (2003). *Teaching for quality learning at university* (2nd ed.). Glasgow, UK: Open University Press.

Blood, E., & Neel, R. (2008). Using student response systems in lecture-based instruction: does it change student engagement and learning? *Journal of Technology and Teacher Education, 16*(3), 375–383.

Bojinova, E. D., & Oigara, J. N. (2011). Teaching and learning with clickers: are clickers good for students? *Interdisciplinary Journal of E-Learning and Learning Objects, 7*, 169–184.

Boyle, J. T., & Nicol, D. J. (2003). Using classroom communication systems to support Interaction and discussion in large class settings. *Association of Learning Technology Journal, 11*(3), 43–57.

Brewer, C. A. (2004). Near real-time assessment of student learning and understanding in biology courses. *BioScience, 54*(11), 1034–1039. http://dx.doi.org/10.1641/0006-3568(2004)054[1034:NRAOSL]2.0.CO;2.

Bruff, D. (2009). *Teaching with classroom response systems: Creating active learning environments*. San Francisco, CA: Jossey-Bass.

Bunce, D. M., VandenPlas, J. R., & Havanki, K. L. (2006). Comparing the effectiveness on student achievement of a student response system versus online WebCT quizzes. *Journal of Chemical Education, 83*(3), 488–493. http://dx.doi.org/10.1021/ed083p488.

Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: does it affect learning from lectures? *Applied Cognitive Psychology, 23*(6), 747–759. http://dx.doi.org/10.1002/acp.1513.

Cardoso, W. (2011). Learning a foreign language with a learner response system: the students' perspective. *Computer Assisted Language Learning, 24*(5), 393–417. http://dx.doi.org/10.1080/09588221.2011.567354.

Carpenter, C. R. (1950a). *The classroom communicator (rapid mass learning)*. Technical report. (P. W. Office of Naval Research, NY. Special Devices Center., Trans.): Pennsylvania State University, University Park. College of Education.

Carpenter, C. R. (1950b). *The film analyzer (rapid mass learning)*. Technical report. (P. W. Office of Naval Research, NY. Special Devices Center., Trans.): Pennsylvania State Univ., University Park. College of Education.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276. http://dx.doi.org/10.1207/s15327906mbr0102_10.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459. http://dx.doi.org/10.3102/00346543053004445.

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

Corcos, E., & Monty, V. (2008). Interactivity in library presentations using a personal response system. *EDUCAUSE Quarterly, 31*(2), 53–60.

Crossgrove, K., & Curran, K. L. (2008). Using clickers in nonmajors- and majors-level biology courses: student opinion, learning, and long-term retention of course material. *CBE-Life Sciences Education, 7*(1), 146–154. http://dx.doi.org/10.1187/cbe.07-08-0060.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: ten years of experience and results. *American Journal of Physics, 69*(9), 970–977. http://dx.doi.org/10.1119/1.1374249.

D'Arcy, Eastburn, & Mullally. (2007). Effective use of a personal response system in a general education plant pathology class. *The Plant Health Instructor*, . http://dx.doi.org/10.1094/PHI-T-2007-0315-07.

DeBourgh, G. A. (2008). Use of classroom "clickers" to promote acquisition of advanced reasoning skills. *Nurse Education in Practice, 8*(2), 76–87. http://dx.doi.org/10.1016/j.nepr.2007.02.002.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: SAGE.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20*(2), 81–94. http://dx.doi.org/10.1111/j.1365-2729.2004.00074.x.

Entwistle, N. (2010). Taking stock: an overview of key research findings. In J. C. Hughes, & J. Mighty (Eds.), *Taking stock: Research on teaching and learning in higher education* (pp. 15–57). Montreal, QC, Canada: McGill-Queen's University Press.

Fagen, A. P., Crouch, C. H., & Mazur, E. (2002). Peer instruction: results from a range of classrooms. *The Physics Teacher, 40*(4), 206–209. http://dx.doi.org/10.1119/1.1474140.

Fava, J. L., & Velicer, W. F. (1992). An empirical comparison of factor, image, component, and scale scores. *Multivariate Behavioral Research, 27*(3), 301–322.

Fies, C., & Marshall, J. (2006). Classroom response systems: a review of the literature. *Journal of Science Education and Technology, 15*(1), 101–109. http://dx.doi.org/10.1007/s10956-006-0360-1.

Fitch, J. L. (2004). Student feedback in the college classroom: a technology solution. *Educational Technology Research and Development, 52*(1), 71–77. http://dx.doi.org/10.1007/BF02504773.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. http://dx.doi.org/10.3102/00346543074001059.

Gauci, S. A., Dantas, A. M., Williams, D. A., & Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education, 33*, 60–71. http://dx.doi.org/10.1152/advan.00109.2007.

George, D., & Mallery, P. (2012). *IBM SPSS statistics 19 step by step: A simple guide and reference* (12th ed.). Boston, MA: Pearson.

Graham, C. R., Tripp, T. R., Seawright, L., & Joeckel, G. (2007). Empowering or compelling reluctant participators using audience response systems. *Active Learning in Higher Education, 8*(3), 233–258. http://dx.doi.org/10.1177/1469787407081885.

Greer, L., & Heaney, P. J. (2004). Real-time analysis of student comprehension: an assessment of electronic student response technology in an introductory earth science course. *Journal of Geoscience Education, 52*(4), 345–351.

Hoekstra, A. (2008). Vibrant student voices: exploring effects of the use of clickers in large college courses. *Learning, Media and Technology, 33*(4), 329–341. http://dx.doi.org/10.1080/17439880802497081.

Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. http://dx.doi.org/10.1007/bf02289447.

James, M. C. (2006). The effect of grading incentive on student discourse in peer instruction. *American Journal of Physics, 74*(8), 689–691. http://dx.doi.org/10.1119/1.2198887.

James, M. C., Barbieri, F., & Garcia, P. (2008). What are they talking about? Lessons learned from a study of peer instruction. *Astronomy Education Review, 7*(1), 37–43. http://dx.doi.org/10.3847/AER2008004.

Johnson, T., & Meckelborg, A. (2008). *Student response systems: A cure for lecturalgia?*. Paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008, Vienna, Austria.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark Iv. *Educational and Psychological Measurement, 34*(1), 111–117. http://dx.doi.org/10.1177/001316447403400115.

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: a review of the literature. *Computers & Education, 53*(3), 819–827. http://dx.doi.org/10.1016/j.compedu.2009.05.001.

King, S. O., & Robinson, C. L. (2009). 'Pretty Lights' and maths! Increasing student engagement and enhancing learning through the use of electronic voting systems. *Computers & Education, 53*(1), 189–199. http://dx.doi.org/10.1016/j.compedu.2009.01.012.

Knight, P., & Yorke, M. (2003). *Assessment learning and employability*. Maidenhead, UK: McGraw-Hill Education.

Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: from Harvard to the two-year college. *American Journal of Physics, 76*(11), 1066–1069. http://dx.doi.org/10.1119/1.2978182.

Latessa, R., & Mouw, D. (2005). Use of an audience response system to augment interactive learning. *Family Medicine, 37*(1), 12–14.

MacGeorge, E., Homan, S., Dunning, J., Elmore, D., Bodie, G., Evans, E., et al. (2008). Student evaluation of audience response technology in large lecture classes. *Educational Technology Research and Development, 56*(2), 125–145. http://dx.doi.org/10.1007/s11423-007-9053-6.

McKeachie, W. J. (1990). Research on college teaching: the historical background. *Journal of Educational Psychology, 82*(2), 189–200. http://dx.doi.org/10.1037/0022-0663.82.2.189.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology, 34*(1), 51–57. http://dx.doi.org/10.1016/j.cedpsych.2008.04.002.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

Miller, R. G., Ashar, B. H., & Getz, K. J. (2003). Evaluation of an audience response system for the continuing education of health professionals. *Journal of Continuing Education in the Health Professions, 23*(2), 109–115. http://dx.doi.org/10.1002/chp.1340230208.

Mollborn, S., & Hoekstra, A. (2010). "A Meeting of Minds": using clickers for critical thinking and discussion in large sociology classes. *Teaching Sociology, 38*(1), 18–27. http://dx.doi.org/10.1177/0092055x09353890.

Moredich, C., & Moore, E. (2007). Engaging students through the use of classroom response systems. *Nurse Educator, 32*(3), 113–116. http://dx.doi.org/10.1097/01.NNE.0000270225.76661.74.

Morling, B., McAuliffe, M., Cohen, L., & DiLorenzo, T. M. (2008). Efficacy of personal response systems ("Clickers") in large, introductory psychology classes. *Teaching of Psychology, 35*(1), 45–50. http://dx.doi.org/10.1080/00986280701818516.

Mulaik, S. A. (2010). *The foundations of factor analysis* (2nd ed.). Boca Raton, FL: CRC Press.

Nicol, D. J., & Boyle, J. T. (2003). Peer instruction versus class-wide discussion in large classes: a comparison of two interaction methods in the wired classroom. *Studies in Higher Education, 28*(4), 457–473. http://dx.doi.org/10.1080/0307507032000122297.

Offerdahl, E. G., & Tomanek, D. (2011). Changes in instructors' assessment thinking related to experimentation with new strategies. *Assessment & Evaluation in Higher Education, 36*(7), 781–795. http://dx.doi.org/10.1080/02602938.2010.488794.

Oliver, B., Tucker, B., Gupta, R., & Yeo, S. (2008). eVALUate: an evaluation instrument for measuring students' perceptions of their engagement and learning outcomes. *Assessment & Evaluation in Higher Education, 33*(6), 619–630. http://dx.doi.org/10.1080/02602930701773034.

Penuel, W., Boscardin, C., Masyn, K., & Crawford, V. (2007). Teaching with student response systems in elementary and secondary education settings: a survey study. *Educational Technology Research and Development, 55*(4), 315–346. http://dx.doi.org/10.1007/s11423-006-9023-4.

Poirier, C. R., & Feldman, R. S. (2007). Promoting active learning using individual response technology in large introductory psychology classes. *Teaching of Psychology, 34*(3), 194–196. http://dx.doi.org/10.1080/00986280701498665.

Pradhan, A., Sparano, D., & Ananth, C. V. (2005). The influence of an audience response system on knowledge retention: an application to resident education. *American Journal of Obstetrics and Gynecology, 193*, 1827–1830. http://dx.doi.org/10.1016/j.ajog.2005.07.075.

Reay, N. W., Li, P., & Bao, L. (2008). Testing a new voting machine question methodology. *American Journal of Physics, 76*(2), 171–178. http://dx.doi.org/10.1119/1.2820392.

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28*(2), 147–169. http://dx.doi.org/10.1023/B: MOEM.0000032312.95499.6f.

Rice, R. E., & Bunz, U. (2006). Evaluating a wireless course feedback system: the role of demographics, expertise, fluency, competency, and usage. *Studies in Media and Information Literacy Education, 6*(3), 1–23. http://dx.doi.org/10.3138/sim.6.3.002.

Roselli, R. J., & Brophy, S. P. (2006). Experiences with formative assessment in engineering classrooms. *Journal of Engineering Education, 95*, 311–324.

Saroyan, A., & Amundsen, C. (Eds.), (2004). *Rethinking teaching in higher education*. Sterling, VA: Stylus.

Shaffer, D. M., & Collura, M. J. (2009). Evaluating the effectiveness of a personal response system in the classroom. *Teaching of Psychology, 36*(4), 273–277. http://dx.doi.org/10.1080/00986280903175749.

Siau, K., Sheng, H., & Nah, F. F. (2006). Use of a classroom response system to enhance classroom interactivity. *IEEE Transactions on Education, 49*(3), 398–403. http://dx.doi.org/10.1109/TE.2006.879802.

Simpson, V., & Oliver, M. (2007). Electronic voting systems for lectures then and now: a comparison of research and practice. *Australasian Journal of Educational Technology, 23*(2), 187–208.

Stein, P. S., Challman, S. D., & Brueckner, J. K. (2006). Using audience response technology for pretest reviews in an undergraduate nursing course. *Journal of Nursing Education, 45*(11), 469–473.

Steinberg, J. (2010, November 16). *More professors give out hand-held devices to monitor students and engage them*. The New York Times.

Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: the state-of-the-art of the research. *Educational Research Review, 5*(1), 25–49. http://dx.doi.org/10.1016/j.edurev.2009.07.001.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Lawrence Erlbaum.

Stevenson, F. (2007). Clickers: the use of audience response questions to enliven lectures and stimulate teamwork. *Journal of International Association of Medical Science Educators, 17*(2), 106–111.

Stowell, J. R., & Nelson, J. M. (2007). Benefits of electronic audience response systems on student participation, learning, and emotion. *Teaching of Psychology, 34*(4), 253–258. http://dx.doi.org/10.1080/00986280701700391.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Upper Saddle River, NJ: Pearson Education.

Trigwell, K. (2010). Teaching and learning: a relational view. In J. C. Hughes, & J. Mighty (Eds.), *Taking stock: Research on teaching and learning in higher education* (pp. 115–128). Montreal, QC, Canada: McGill-Queen's University Press.

Van Dijk, L. A., Van Den Berg, G. C., & Van Keulen, H. (2001). Interactive lectures in engineering education. *European Journal of Engineering Education, 26*(1), 15–28. http://dx.doi.org/10.1080/03043790123124.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*(1), 1–28.

Weimer, M., & Lenze, L. F. (1991). Instructional interventions: a review of the literature and interventions. In J. C. Smart (Ed.). *Higher education: Handbook of theory and research, Vol. VII* (pp. 294–333). New York, NY: Agathon Press.

Weston, C., & Cranton, P. A. (1986). Selecting instructional strategies. *The Journal of Higher Education, 57*(3), 259–288.

Weston, C., Le Maistre, C., McAlpine, L., & Bordonaro, T. (1997). The influence of participants in formative evaluation on the improvement of learning from written instructional materials. *Instructional Science, 25*(5), 369–386. http://dx.doi.org/10.1023/a:1003073923764.

Wieman, C. (2010). Why not try a scientific approaches to science education? In J. C. Hughes, & J. Mighty (Eds.), *Taking stock: Research on teaching and learning in higher education* (pp. 175–190) Montreal, QC, Canada: McGill-Queen's University Press.

Williams, B., & Boyle, M. (2008). The use of interactive wireless keypads for interprofessional learning experiences by undergraduate emergency health students. *International Journal of Education and Development Using ICT, 4*(1), Available at: http://ijedict.dec.uwi.edu/viewarticle.php?id=405.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: SAGE.