



ELSEVIER

Contents lists available at ScienceDirect

Educational Research Review

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

Review

Do we click in the right way? A meta-analytic review of clicker-integrated instruction

Yu-Ta Chien ^{a, b}, Yueh-Hsia Chang ^c, Chun-Yen Chang ^{a, b, d, *}^a Graduate Institute of Science Education, National Taiwan Normal University, Taiwan^b Science Education Center, National Taiwan Normal University, Taiwan^c Graduate Institute of Curriculum and Instruction, Tamkang University, Taiwan^d Department of Earth Sciences, National Taiwan Normal University, Taiwan

ARTICLE INFO

Article history:

Received 26 November 2014

Received in revised form 13 October 2015

Accepted 13 October 2015

Available online 23 October 2015

Keywords:

Clicker

Instant response system

Clicker-integrated instruction

Learning outcome

Academic learning

ABSTRACT

Clickers, formerly known as instant response systems, have gradually become an integral part of the classroom. Though several reviews on research into clicker-integrated instruction have been published within this decade, the controversy over whether clicker-integrated instruction is effective to enhance students' learning gains has not been settled because the early reviews mainly focus on students' perceptions toward and acceptance of clicker-integrated instruction. Furthermore, so far there is no consistent and clear framework to explain why the use of clickers is effective or ineffective to facilitate academic learning outcomes. Based on the literature from the 1970s to the early 2010s, this review article identifies and summarizes the theoretical aspects accounting for possible relations between clicker-integrated instruction and academic learning outcomes. The theoretical aspects are subsequently evaluated and expanded in reference to primary studies. The results suggest that the superior effect of clicker-integrated instruction, compared to conventional lectures, stands on firm empirical ground. In addition, engaging students in explaining and justifying their answers to clicker questions is highly recommended because such an instructional strategy is associated with positive and strong effect sizes on academic learning outcomes.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1.	Introduction	2
2.	Strategies and theoretical aspects of implementing clickers into the classroom	3
2.1.	Novelty effect	4
2.2.	Unequal-item exposure effect	4
2.3.	Testing effect	4
2.4.	Adjunct-question effect	4
2.5.	Feedback-intervention effect	5
2.6.	(Self-)explanation effect	5
2.7.	Summary	6

* Corresponding author. Graduate Institute of Science Education, Department of Earth Sciences, Science Education Center, National Taiwan Normal University, 88, Section 4, Ting-Chou Road, Taipei 11677, Taiwan.

E-mail address: changcy@ntnu.edu.tw (C.-Y. Chang).

<http://dx.doi.org/10.1016/j.edurev.2015.10.003>

1747-938X/© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

3.	Method	6
3.1.	Selection of studies	6
3.2.	Coding of study features	8
3.3.	Calculation and analysis of effect sizes	8
4.	Results	9
4.1.	Main characteristics of selected studies	9
4.2.	General outcomes of clicker-integrated instruction	9
4.3.	Relationship between instruction length and learning outcomes	12
4.4.	Relationship between question-answering activities and learning outcomes	12
4.5.	Relationship between peer discussion and learning outcomes	12
5.	Discussion	14
6.	Conclusion	15
	Acknowledgements	16
	References	16

1. Introduction

Clickers, formerly known as Instant Response Systems (IRS), have gradually become an integral part of the classroom. According to a survey by CNET News (Gilbert, 2005), in 2004 alone, schools and universities around the world bought nearly a million clickers. Moreover, it was estimated that 8 million clickers were sold annually by 2008. Based on these statistics, it might be said that tens of thousands of courses worldwide are now being conducted with the addition of clickers. As indicated by several reviews (e.g., Boscardin & Penuel, 2012; Caldwell, 2007; Fies & Marshall, 2006; MacArthur & Jones, 2008), the instructional potential of clickers has grabbed substantial attention from researchers and educators in various disciplines.

By summarizing the information documented in the early reviews of clicker use (e.g., Boscardin & Penuel, 2012; Caldwell, 2007; Fies & Marshall, 2006; MacArthur & Jones, 2008), it can be found that clickers are signal transmitters, similar in size to television remotes, used to collect students' responses to teachers' questions in the classroom. Once the teacher poses a question, generally a multiple-choice type inquiry, students can click the buttons on their remote-like devices to specify answers to that question. Students' answers are then transmitted to a monitoring system typically through infrared or radio frequency signals. By this means, every student in the classroom can, hopefully, express his/her thoughts instantly without being scrutinized by peers. The monitoring system then automatically aggregates the answers from the entire class with a histogram, offering the teacher a choice about whether or not the overall distribution of students' answers should be publicly shown. Some systems likewise enable teachers to decide how publicly or how anonymously students' responses are collected and displayed by indicating which remotes have transmitted signals or by the names registered for the remotes.

Regarding the widespread use of clickers in schools, not surprisingly, several researchers and educators have questioned its cost and effectiveness. For instance, Lasry (2008) quoted Marvin Davis, "as men get older, the toys get more expensive," to express his concern on the benefits of clickers that may be overstated by companies of educational technology and other researchers. Lantz (2010) also pointed out the doubt of educators who hesitate to adopt clickers in their own classrooms: are clickers a worthwhile, pedagogical tool or merely an amusing novelty? In order to judge whether clickers are worth the investment, the following questions are raised: Can clickers be used to complement the existing practice in school, or even to facilitate new instructional strategies that are difficult, if not impossible, to achieve in a conventional classroom setting? These questions should be carefully examined in reference to how students learn and what role this new technology plays in the learning processes. Reviews on the empirical studies into clicker-integrated instruction are certainly helpful to answer these questions.

Several reviews on research into clicker-integrated instruction have been published within this decade (e.g., Caldwell, 2007; Fies & Marshall, 2006; Kay & LeSage, 2009; Lantz, 2010; MacArthur & Jones, 2008; Simpson & Oliver, 2007). However, the controversy over whether clicker-integrated instruction is effective to enhance students' learning gains has not been settled because, as indicated by Fies and Marshall (2006) and Kay and LeSage (2009), little empirical work was done in the early 2000's. Furthermore, the most frequently used data-collection method of empirical studies is self-reported measures (Kay & LeSage, 2009). The major accomplishments of early reviews are the syntheses of why clickers might be worthy to try in the classroom and what problems may occur while implementing clickers into schools (see Caldwell, 2007; Fies & Marshall, 2006; Kay & LeSage, 2009; Lantz, 2010; MacArthur & Jones, 2008; Simpson & Oliver, 2007, for reviews). Two solid conclusions can be drawn from these syntheses: students usually hold positive attitudes toward the use of clickers, and clickers can boost student attendance in higher education. However, a clear explanation of why the use of clickers is effective to facilitate academic learning outcomes remains absent. Several authors of the early reviews (e.g., Boscardin & Penuel, 2012; Fies & Marshall, 2006; Kay & LeSage, 2009) have called for more empirical studies, which should be conducted in a more rigorous manner, to provide explanations for academic learning outcomes through explicit incorporation of a theoretical framework as well as the features of clicker-integrated instruction.

As evidenced by the statistics reported in Section 4.1, a substantial amount of empirical studies have been published in peer-reviewed journals in recent years. A synthesis for these empirical studies is thus needed to update the state of the field.

Therefore, this article presents the results of a meta-analysis on the empirical studies into clicker-integrated instruction. Moreover, theoretical aspects accounting for possible effects of clicker-integrated instruction on academic learning outcomes are reviewed and identified from literature. These theoretical aspects are subsequently evaluated and expanded in reference to the primary studies reviewed in this article. The main questions guided this review are:

- What are the theoretical aspects accounting for possible effects of clicker-integrated instruction on academic learning outcomes?
- Have primary studies demonstrated that clicker-integrated instruction is generally superior to conventional lectures for enhancing academic learning outcomes?
- Which theoretical argument is more aligned with the available empirical data?
- What can we learn from empirical studies to improve our practices with clickers in both teaching and research?

2. Strategies and theoretical aspects of implementing clickers into the classroom

Before proceeding with the theoretical aspects to explain why using clickers is effective to facilitate academic learning outcomes, the features of clicker-integrated instruction, especially, which are different from conventional lectures, should be identified. As shown in Fig. 1, three typical forms of clicker-integrated instruction are summarized from previous literature reviews (e.g., Caldwell, 2007; Fies & Marshall, 2006; Kay & LeSage, 2009; Lantz, 2010; MacArthur & Jones, 2008). It can be found that the instruction is divided into several units, using clicker questions as the central component; within each unit, a clicker question, usually presented in a form of multiple-choice, is posed to students after a brief lecture. The most basic way to proceed with the instruction is to ask students to vote for possible answers individually, followed by a display of the voting results. The instructor then provides students with explanations for correct and incorrect answers. A slight variation from the basic form is to encourage students to discuss with their neighbors before submitting answers. The third form of clicker-integrated instruction, primarily based on Mazur's peer instruction approach (Crouch & Mazur, 2001; Mazur, 1997), is more sophisticated. Once the clicker question is posed, students are asked to vote on answers individually. The display of the

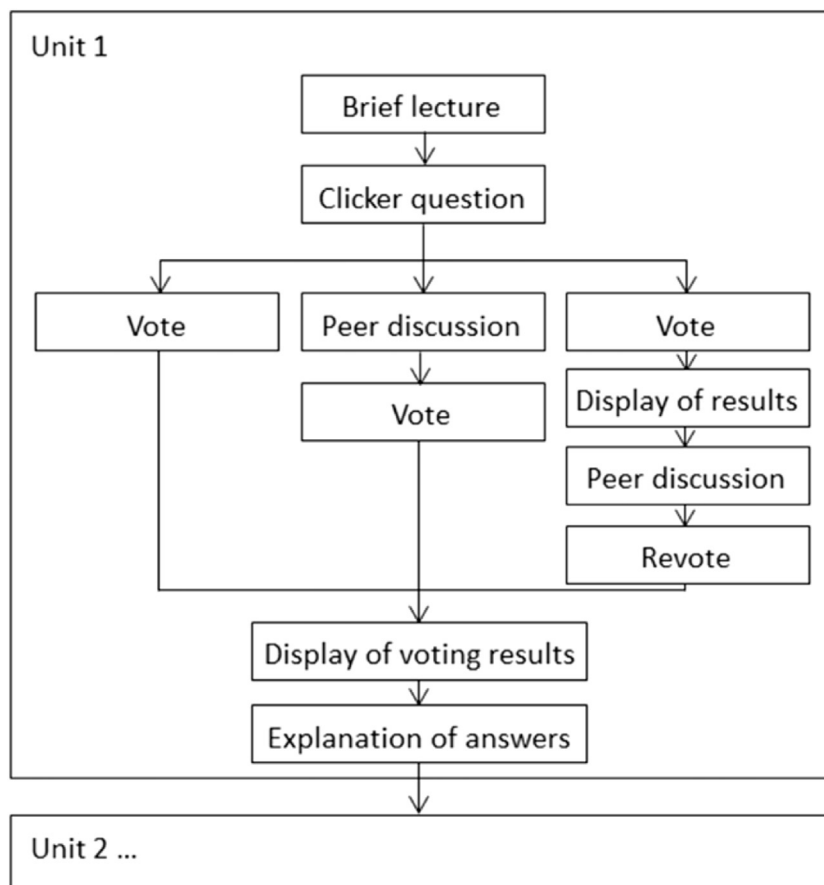


Fig. 1. Typical forms of clicker-integrated instruction.

voting results is then shown, but no explanation for answers is given by the instructor. Rather, by initiating peer discussions, students are engaged in generating their own explanations to justify each choice. After that, students are given a chance to revote on answers. The display, of the revote results, is shown subsequently. The instructor then explains why the answers are correct or incorrect.

As mentioned earlier, so far there is no consistent and clear framework to explain why the use of clickers is effective to facilitate learning outcomes. Nonetheless, several candidate theories can be derived from the research of educational technology and cognitive psychology. In the following sections, these candidates are discussed in conjunction with the features of clicker-integrated instruction. The research designs corresponding to these candidates are described as well.

2.1. Novelty effect

The most straightforward explanation of how research papers produce positive results of clicker-integrated instruction is the *novelty effect*. It assumes that students are often excited when a new technology is adopted in the classroom. The increased attention by students thus results in increased effort or persistence in learning, which yields achievement gains. In this case, the positive effect of that technology on learning outcomes is mainly due to its novelty. Learning gains thus tend to diminish significantly as students become more familiar with the technology (Bangert-Drowns, Kulik, & Kulik, 1985; Cheung & Slavin, 2013; Kulik & Kulik, 1991; Kulik, Kulik, & Bangert-Drowns, 1985). This argument is commonly used to counter the promotion of new technology in academic research (e.g., Clark, 1983, 1994) because no pedagogical benefit is substantially offered by the adoption of the technology.

2.2. Unequal-item exposure effect

As warned by Anthis (2011), the positive results of clicker-integrated instruction may be merely caused by *unequal exposure* to test items between experimental (i.e., clicker-integrated instruction) and control (i.e., conventional lectures) groups. Some studies simply lump the questions posed in clicker-integrated instruction together as a post-test to determine students' learning outcomes. However, the control group does not receive any of the questions during lectures. In other words, the content of the post-test is relatively new to the control group. Under this circumstance, it is premature to claim positive effects of clickers on reforming conventional lectures even if a higher mean score of the experimental group is observed. Shapiro and Gordon (2012) also expressed a similar concern that the learning gains from clicker-integrated instruction may merely come about by prompting students to memorize specific test items during class. If this sort of unequal exposure is at the root of the positive results observed in some clicker studies, the effect is not particularly interesting from a theoretical point of view because it might be just as effective to give students lists of important topics to attend to in class and during study.

2.3. Testing effect

Given that clicker-integrated instruction is inherently a series of test events, the *testing effect* is cited by previous studies to explain the effectiveness of clicker-integrated instruction (e.g., Campbell & Mayer, 2009; Mayer et al., 2009; Shapiro & Gordon, 2012). According to Roediger and Karpicke's reviews (2006a, 2006b) on testing-effect experiments, compared to restudying learning material, taking a test of that material can have a greater positive effect on memory retention and organization. The act of taking tests induces students to retrieve information from their long-term memory. Retrieval may increase the elaboration of a memory trace and multiply retrieval routes, thereby increasing the probability of successful retrieval in the future (Dempster, 1996; Roediger III & Karpicke, 2006a, 2006b). Moreover, activation of a targeted concept in memory may produce facilitative effects for its related concepts because the associative strength between these concepts in memory is increased through retrieving one of them (Chan, 2010; Chan, McDermott, & Roediger III, 2006). The inference drawn from such research design is unlikely to be threatened by unequal-item exposures since the initial tests are usually blank, simply asking students to recall what they have read.

The testing effect has great implications for lecture-based instruction: test your students after you lectured them, or they will forget what you said soon after, even when you have repeated the lecture several times. Clickers can facilitate in-class tests with ease. However, the clicker questions posed in the classroom usually are of higher cognitive levels, instead of free-recall. Furthermore, clicker questions are usually given at multiple time points, rather than massed together at the end of the class. Does the testing effect remain under these circumstances? Other studies from a different research tradition are reviewed to deal with these issues in the next section.

2.4. Adjunct-question effect

The *adjunct-question effect* is cited in some primary studies to account for the effectiveness of clicker-integrated instruction (e.g., Campbell & Mayer, 2009; Mayer et al., 2009). Based on the comprehensive reviews by Hamilton (1985) and Rickards (1979), adjunct questions refer to the questions inserted in a written passage that students are to study. Literally, the design of adjunct-question experiments is pretty much alike that of testing-effect experiments; they both use test-events without feedback to facilitate learning outcomes. Nonetheless, some variations between them can be identified. One of

the variations is that adjunct questions are typically interspersed into reading material at a number of points, whereas the questions used in testing-effect experiments are placed together at the end of the material. In addition, a number of adjunct-question experiments have used higher-order questions that go beyond just asking students to recognize, recall, or supply some factual information given in reading material. According to the summary by Andre (1979), higher-order adjunct questions require students to select a new example of a concept or principle employed in the material from among alternatives; to state a relationship between elements implied but not explicitly stated in the material; or to perform cognitive tasks, which are more complex than the knowledge (remembering) level of the Bloom Taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). The meta-analysis by Hamaker (1986) reveals that, compared to restudying learning material, answering adjunct questions are generally more effective to promote students' learning gains.

Two explanations for the adjunct-question effect are identified from relevant literature (e.g., Frase, 1967; Hamaker, 1986; Rickards, 1979). The first one is backward processing, which is similar to the mechanism of retrieval-induced facilitation, referring to the act of retrieving information thematically related to the adjunct question from memory. Another explanation is forward processing, referring to increased attention to the instructional material immediately following the adjunct question. In terms of clicker-integrated instruction, these two explanations should not be seen inclusive of each other. In a practical sense, the clicker questions posed in a lecture should be highly inter-related. Both a backward and a forward process might thus be produced by the clicker questions. In summary, the features of adjunct-question experiments are more aligned with clicker-integrated instruction than those of testing-effect experiments.

2.5. Feedback-intervention effect

In clicker-integrated instruction, students may perceive two kinds of information as feedback, regarding their performance on clicker questions. As depicted in Fig. 1, the first one is the display of voting results. The second one is teachers' explanations for the answers to clicker questions. Feedback is important for cueing students how to improve their learning performance (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Butler & Winne, 1995; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008; Thurlings, Vermeulen, Bastiaens, & Stijnen, 2013). Based on a large meta-analysis of empirical studies, Kluger and DeNisi (1996) polarize the attunement in locus of attention induced by feedback: at one end, are the details of the focal task, whereas, at the other end are the threats to self. The effectiveness of feedback decreases as a student's attention moves closer to the self and away from the task. Hattie and Timperley (2007) propose a more concrete model to predict feedback effects. The model addresses feedback with reference to the process of learning across four levels, including: (1) the *task* level, indicating how well tasks are understood or performed; (2) the *process* level, referring to the main process needed to understand or perform tasks; (3) the *self-regulation* level, focusing on the ways in which students monitor, direct, and regulate themselves toward learning goals; and (4) the *self* level, inducing personal evaluations about the self (Hattie & Timperley, 2007). Effective feedback should direct students' attention to how to improve task performance (i.e., the task, process, and self-regulation levels), rather than personal evaluations about the self.

A possible strength of clickers is to generate frequent performance-feedback-adjustment loops for every student in a single lecture. If feedback is delayed or absent when a student struggles with learning material for a while, his/her attention may shift from the focal task to the self – self-doubting whether he/she is incapable of completing the task (Kluger & DeNisi, 1996; Shute, 2008). The student may therefore feel too frustrated to keep learning. As shown in Fig. 1, clicker-integrated instruction usually chunks the learning material into smaller units; one or a few clicker questions follow each unit to engage students and track their progress. Feedback, referring to either the display of voting results or teachers' explanations, is then offered. Students may adapt feedback to evaluate and adjust the gap between their current status and the learning goal. The performance-feedback-adjustment loop is difficult to iteratively implement in the classroom without an efficient and reliable tool for data collection and calculation. Clickers are helpful at this point. Recent studies also have suggested that the performance-feedback-adjustment loop can facilitate students' improvement in metacognitive awareness, which helps students to interpret feedback at the self-regulation level (e.g., Azevedo & Aleven, 2013; Goldberg & Spain, 2014).

The privacy enabled by clickers might be helpful for preventing students' attention from being misdirected to the self level. In a traditional classroom, the teacher usually asks students to exercise academic tasks in a public manner, such as performing tasks at the podium or raising hands to express opinions. This instructional strategy is risky because it exposes students' weakness to peers and thus may evoke their fears of failure. Students may then interpret feedback from the teacher and peers as threats to the self if they feel their own performance is poor, wrong, or unpopular. Under such the circumstance, feedback deflects students' attention away from how to improve task performance, and usually has a negative effect on learning (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). As reported in Hoekstra's observational study (2008), expressing opinions through clickers is less anxiety provoking than traditional methods (e.g. raising hands or holding up a response card that peers can see). Thus students may interpret feedback in positive ways that attain task, process, or self-regulation levels.

2.6. (Self-)explanation effect

Considering that some clicker-integrated instruction asks students to explain their reasoning on clicker questions to peers, the research on the *self-explanation effect* may provide theoretical underpinnings for such an instructional strategy. The possible causal relationship, between academic learning outcomes and the behavior of generating explanations for problem solutions while learning, was initially investigated by Chi and her colleagues (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi,

De Leeuw, Chiu, & Lavancher, 1994; Chi & VanLehn, 1991). It was found that prompting students to self-explain, while reading prescribed step-by-step problem solutions, was more effective in facilitating conceptual learning than asking students to read the problem solutions twice without prompts (Chi et al., 1994). Since then, a substantial amount of empirical evidence was obtained to demonstrate that prompting student to explain solutions of worked examples to themselves had a positive effect on their future problem-solving performance (e.g., Bielaczyc, Pirolli, & Brown, 1995; Renkl, Stark, Gruber, & Mandl, 1998; Rittle-Johnson, 2006; Wong, Lawson, & Keeves, 2002). Neuman and Schwarz (1998) further investigated whether the self-explanation effect could be applied to the context in which students were solving problems by themselves, rather merely studying prescribed solutions from examples. The transferability of the self-explanation effect was demonstrated in Neuman and Schwarz's study (1998). Moreover, it was found that explicitly prompting students to generate explanations, regarding the difficulties they had encountered and the corresponding solution steps they had operated, was more effective than asking students to just freely describe what thoughts passed in their heads while solving problems.

Encouraging students to explain their own reasoning on clicker questions can be seen as a prompt for self-explanations. From the cognitive perspective, learning is seen as the process in which the student mentally manipulates his/her own knowledge in conjunction with new information to create newer knowledge (Chi et al., 1989, 1994; Chi & VanLehn, 1991). Engaging the student in self-explaining can facilitate him/her to retrieve, integrate, and modify his/her own knowledge with new information in a minute and ongoing fashion. New declarative or procedural knowledge, which can subsequently be used during problem solving, may thus be better constructed by the student (Chi et al., 1989, 1994; Chi & VanLehn, 1991).

The merits of self-explaining have become more recognized by contemporary educational researchers from the metacognitive perspective. As widely observed in empirical studies (Bielaczyc et al., 1995; Chi et al., 1989; Chi & VanLehn, 1991; Renkl, 1997; Wong et al., 2002), successful students generally tend to generate more task-related self-explanations while solving problems, than poor students. Moreover, the content of self-explanations is partially related to how they are going to monitor and regulate their goals, strategies, and performance. This implies that self-explaining may be a strategy for students to maintain metacognitive awareness. By generating self-explanations, students are more likely to find gaps in their knowledge that are causing them trouble, and then invent as small a piece of knowledge as necessary for filling the gaps (Bielaczyc et al., 1995; Chi & Bassok, 1989; Renkl et al., 1998; VanLehn, Jones, & Chi, 1992; Wong et al., 2002). Recent studies in educational technology have also indicated that self-explanation prompts can promote students' metacognitive processes, of detecting and correcting errors, and thus facilitate students in acquiring a deeper understanding of learning materials (Alevin & Koedinger, 2002; Atkinson, Renkl, & Merrill, 2003; Azevedo & Alevin, 2013; Chi & VanLehn, 2010; Goldberg & Spain, 2014; Graesser, McNamara, & VanLehn, 2005; Mathan & Koedinger, 2005). The possible effects of asking students to explain their reasoning on clicker questions, and its mechanism, should be also considered from social aspects. As shown in Fig. 1, students are asked to explain their reasoning to peers, rather than merely to themselves. In such the setting, social interactions occur. Students might use peers' actions and utterances as external information to monitor and regulate their cognition, motivation, and behavior with reference to the group he/she is involved with (Azevedo & Alevin, 2013; Järvelä et al., 2015; Kirschner, Kreijns, Phielix, & Franssen, 2015; Kreijns, Kirschner, & Vermeulen, 2013; Phielix, Prins, & Kirschner, 2010; Phielix, Prins, Kirschner, Erkens, & Jaspers, 2011).

2.7. Summary

The theoretical aspects discussed above are summarized in Table 1. Moreover, the features of research design which are critical to support each of the theoretical aspects are specified, including (1) whether instruction is conducted with a relatively longer duration; (2) whether a new set of questions are used to construct post-tests; (3) whether the control group receives in-class questions as does the experimental group; (4) whether feedback is given to students after they respond to in-class questions; and (5) whether students are explicitly asked to elaborate and justify their answers (e.g., by peer discussion). For each theoretical aspect, the relevant features are marked in the row of that theoretical aspect. The predictions of research results are described as well. Acknowledging that many of the studies discussed above are not directly focused on clicker-integrated instruction, these studies alone cannot fully inform on educational practice of clicker-integrated instruction. Therefore, a meta-analytic review on primary research in clicker-integrated instruction is conducted to evaluate the appropriateness of the aforementioned theoretical aspects. The meta-analysis is subsequently guided by the summary listed in Table 1.

3. Method

3.1. Selection of studies

Several online databases were employed to search relevant studies written in English, including *ERIC*, *JSTOR*, *PsycINFO*, *PsyARTICLE*, *PubMed*, *SCI*, and *SSCI*. Various keywords such as *audience response system*, *classroom response system*, *wireless response system*, *electronic voting system*, *group response system*, *personal response system*, *instant response system*, *interactive voting system*, *student response system*, *clicker*, etc. were used. In order to enhance the rigidity of data collection, the search was limited to research papers published by peer-reviewed journals. The search was conducted at the end of 2013, and was not limited to a particular date range. A total of 556 non-overlapping research papers were located. The large number of hits was screened on the basis of their abstracts to decide whether the obtained articles should be included in the meta-analysis. Once

Table 1
 Descriptions, features, and predictions of relevant theoretical aspects.

Theoretical aspect	Description	Key variable					Prediction
		Extending the duration of instruction	Using a new set of questions as a post-test	Administrating in-class questions to the control group	Providing students with feedback	Explicitly asking students to elaborate and justify their answers (e.g., peer discussion)	
Novelty effect	The novelty of clickers appears to students. The increased attention by students thus results in increased effort or persistence in learning, which yields achievement gains.	✓					The positive effect will dissipate significantly if the duration of instruction is extended.
Unequal-item exposure effect	Students in the clicker group have more chances to practice post-test items than those in the control group. Repeated practice results in superior performance.		✓	✓			The positive effect will dissipate significantly if students' chances to practice post-test items are equalized.
Testing effect	Answering a click question induces students to retrieve information relevant to the concepts learned before. Retrieval may increase the elaboration of a memory trace and multiply retrieval routes, thereby increasing the probability of successful retrieval in the future. Activation of a targeted concept in memory may also increase its associative strength with related concepts. The probability of successful retrieval of these concepts in the future is thus increased.		✓				Students' learning achievements will be improved if a number of clicker questions are given in the class.
Adjunct-question effect	Answering a click question induces students to retrieve information thematically related to the concepts learned before. It may also increase students' attention to the instructional materials immediately following the click question.		✓				Students' learning achievements will be improved if a number of clicker questions are interspersed in the class.
Feedback-intervention effect	Feedback alters the locus of attention. It triggers proper cognitive processes (e.g., to encode correct information, which is relevant to instructional materials).		✓	✓	✓		Students' learning achievements will be improved if feedback of clicker questions is given.
(Self-) Explanation effect	Generating explanations while solving a problem supports the localizing and filling of knowledge gaps. It thus facilitates the integration of new learned materials with existing knowledge.		✓	✓		✓	Students' learning achievements will be improved if they are engaged in elaborating and justifying the answers of clicker questions.

a search result lacked the abstract portion, or sufficient information for making a confidential judgment, its full document was retrieved and examined. Any article in this stage would be retained if it (1) used clickers for educational purposes, not for making a clicking sound or training animals; (2) conducted empirical studies to examine the instructional effectiveness of clicker-integrated instruction; (3) involved at least one conventional lecture as the control group; (4) reported on academic learning outcomes; and (5) documented quantitative data of academic test scores.

The ancestry approach was also used to exhaustively search for the peer-reviewed research papers relevant to the present analysis; additional articles were located by checking reference lists of early reviews (e.g., [Boscardin & Penuel, 2012](#); [Caldwell, 2007](#); [Fies & Marshall, 2006](#); [Kay & LeSage, 2009](#); [Lantz, 2010](#); [MacArthur & Jones, 2008](#); [Nelson, Hartling, Campbell, & Oswald, 2012](#)). These articles were also filtered by the inclusion criteria which were mentioned in the previous paragraph. A research assistant was instructed to examine the search results independently. Any doubt in the inclusion of articles was resolved by discussion. The inter-rater reliability was good, estimated by $\kappa = 0.82$. The full documents of the retained articles were then carefully inspected. In this stage, the studies without specifying the basic statistics needed for computing effect sizes, especially sample sizes, were excluded. In the end, 72 pair-wise comparisons, derived from 28 articles (marked with an asterisk in References), were included in the meta-analysis.

3.2. Coding of study features

The design features of each pair-wise comparison, which may help to examine the instructional strategies and theoretical underpinnings of clicker-integrated instruction, were identified. This job was done by two educational researchers independently, consistently following the coding scheme shown in [Table 2](#). There was a substantial agreement between the researchers, estimated by $\kappa = 0.76$. Any doubt in coding results was resolved by discussion. Biographical information of the selected studies was also listed, including the year of publication, sample sizes, instructional domains, and sample characteristics (e.g., graduates, undergraduates, high-schools, etc.).

3.3. Calculation and analysis of effect sizes

Meta-analysis is a research method which combines and summarizes quantitative information from different studies. Researchers thus are enabled to investigate, systematically, relationships between variables that otherwise would be undetectable or difficult to verify in a single study. While different studies assess the same outcome (e.g., the academic learning outcome targeted by this meta-analysis), they may measure it in a variety of ways (i.e., using different tests). Making the quantitative data from different studies comparable is therefore an important task. A common practice in social sciences to deal with this task over past decades is to calculate the Cohen's d coefficient ([Cohen, 1988](#)) for each pair-wise comparison. Cohen's d for any comparison is defined as the difference between the means of two groups (e.g., experimental and control groups) divided by the pooled standard deviation of the two groups (see [Eq. \(1\)](#)). A d coefficient of 0.2 indicates that one-fifth standard deviation separates two means, whereas a d coefficient of 0.5 represents half of a standard deviation unit. This calculation provides researchers with a common language to express the possible relationship between variables in a uniform scale. The magnitudes of possible relationships between different interventions and outcome measures are standardized, with reference to the variability observed in each study. By using the standardized difference in means, commonly termed as the standardized effect size, results of different primary studies can be combined, compared, and summarized in a systematic way.

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1 + n_2 - 2}}} \quad (1)$$

where \bar{X}_i , S_i , and n_i are the sample mean, standard deviation, and size of group i .

It should be noted that, in [Eq. \(1\)](#), pooling the two sample estimates of standard deviations is for obtaining a more accurate estimate of their common value; due to the sampling issues, the sample estimates S_1 and S_2 are unlikely to be identical even if

Table 2
Coding scheme of study features.

Feature	Description	Code
Delayed testing	Was the learning outcome assessed by a delayed post-test?	Yes (Y); No (N)
Baseline control	Were prior differences between groups (i.e., prior knowledge) were controlled by randomly assigning subjects to groups, systematically assigning subjects to groups, or using pre-test scores as covariates?	Yes (Y); No (N)
One-shot	Did the instruction sustain only one-lecture long (typically around 50 min) or less?	Yes (Y); No (N)
Equivalent exposure	Did the control group receive in-class questions as did the experimental group?	Yes (Y); No (N)
Repeated questions	Did the post-test simply repeat in-class questions?	Yes (Y); No (N)
Peer discussion	Was the experimental group allowed to discuss with peers while answering in-class questions?	Yes (Y); No (N)
Display	Were the voting results shown to subjects during instruction?	Yes (Y); No (N)
Elaboration	Were subjects provided with explanations of the correct/incorrect answers for in-class questions?	Yes (Y); No (N)

researchers assume that the underlying population standard deviations of the two groups are the same. However, Hedges and colleagues (Hedges, 1981; Hedges & Olkin, 1985) have indicated that the population variance will be downwardly biased when it is estimated by the pooled standard deviation. It turns out that the effect size will be overestimated especially when sample sizes are small. Nevertheless, this bias can be approximately corrected through multiplication by a factor proposed by Hedges (1981, 1982). So, the corrected effect size nowadays is usually called Hedges' g (Hedges & Olkin, 1985). The common expression of this correction is shown in Eq. (2). Hedges' g provides a superior estimate of the standardized difference in means with small samples. On the other hand, as the sample size increases, Hedges' g and Cohen's d will converge to the same value. In this sense, the magnitude of Hedges' g can be interpreted in Cohen's conventions (e.g., Cohen, 1988; Fritz, Morris, & Richler, 2012; Kirk, 1996); g coefficients of 0.2, 0.5, and 0.8 can be seen as small, medium, and large, respectively. The use of g , rather than d , has been gradually recommended and adopted in recent years because educational researchers become more aware of the fact that the sample sizes involved in educational research are usually small (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009; de Boer, Donker, & van der Werf, 2014; Donker, de Boer, Kostons, Dignath van Ewijk, & van der Werf, 2014; Fritz et al., 2012; Scammacca, Roberts, & Stuebing, 2014).

$$g = \left[1 - \frac{3}{4(n_1 + n_2) - 9} \right] d \quad (2)$$

The standardized effect sizes presented in this meta-analysis were calculated using Hedges' g . It was noted that several pair-wise comparisons were inter-dependent because their samples totally or partially overlapped with each other. The shifting unit of analysis approach (Cooper, 1982, 1998) was employed to retain the information from each pair-wise comparison as much as possible while preserving statistical independence. In this meta-analysis, all pair-wise comparisons were sorted by the type of outcome measures, depending on whether a delayed post-test was used. Within this categorization, moderator analyses were performed. For each moderator analysis, inter-dependent effect sizes were aggregated based upon the particular moderator variable (such as baseline control, one-shot, equivalent exposure, and so forth, as specified in Table 2). Although this strategic compromise could not totally eliminate the problem of independence, it did minimize violations of assumptions about the independence of effect sizes, whilst preserving as much of the data as possible (Cooper, 1998). While combining inter-dependent effect sizes, Borenstein et al.'s (2009) approach was used to balance the weights of multiple effect sizes that were obtained from one single study. For instance, when a study reported two inter-dependent comparisons, a combined effect size was calculated based on the relative weights of these two comparisons. The combined effect size gave more weight to the comparison with a larger sample size than to another comparison in the same study with a smaller sample size. Moreover, the variance of this combined effect size was computed in a manner that took into account the proportion of the repeated sample across comparisons.

After that the dependence issues were intentionally controlled, the random-effects model (Borenstein et al., 2009; Hedges & Vevea, 1998; Raudenbush, 2009; Shadish & Haddock, 2009), rather than the fixed-effect model, was used to calculate the mean effect of each variable of interest. This decision was made due to the following reasons. First, the random-effects model aimed to estimate the mean of a distribution of true effects, rather than a single true effect. It assumed that each study was estimating an effect size for its unique population. Such an assumption was more practical than that of the fixed-effect model because there was no valid reason to believe that the subjects and interventions were exactly the same across the selected studies. The results based on the random-effects model would thus be more generalizable. Second, compared to the fixed-effect model, the random-effects model could assign more balanced weights to studies because it considered both within- and between-studies variances. Therefore, the weights of the studies with large samples would not be disproportionately inflated, whereas the studies with small samples would not be totally ignored. The coded information and effect sizes of the selected primary studies can be found in Table 3. The analysis was done primarily using the Comprehensive Meta-Analysis software version 2.2.

4. Results

4.1. Main characteristics of selected studies

As shown in Table 4, it is very clear that the spreading speed of clickers is much faster than the production speed of empirical studies; before 2008, few studies were done to empirically compare the instructional effectiveness of clicker-integrated instruction with conventional lectures. Most of the pair-wise comparisons (93%) were reported after 2007. Moreover, almost all of these comparisons (93%) used undergraduates or graduates as research samples. Approximately, half of the comparisons (58%) were conducted in the fields of Science, Technology, Engineering, and Mathematics (STEM) education. The homogeneity in sample selection and learning topics would be good, logically, for estimating effects sizes more precisely and reliably.

4.2. General outcomes of clicker-integrated instruction

As shown in Table 4, sixteen (22%) of the comparisons did not use any technique (e.g., randomly or systematically assigning subjects to groups, or using pre-test scores as covariates) to control subjects' prior differences in academic performance

Table 3
Coded information of selected studies.

Study	n_{total}	$n_{pairwise}$	g_{raw}	SE_{raw}	Delayed testing	Baseline control	One-shot	Equivalent exposure	Repeated questions	Peer discussion	Display	Elaboration
Agbatogun, 2012	67	67	1.73	0.29	N	Y	N	N	N	Y	Y	Y
Bachman & Bachman, 2011	209	209	0.27	0.14	N	N	N	N/A	N	N	Y	Y
		192	0.68	0.15	N	N	N	N/A	N	N	Y	Y
Bartsch & Murphy, 2011	52	52	0.57	0.28	N	Y	Y	Y	N	N	Y	N
Butler, Pyzdrowski, Walker, & Yoho, 2010	406	276	-0.19	0.12	N	N	N	Y	N	N	Y	Y
		348	-0.06	0.11	N	N	N	Y	N	N	Y	Y
		406	-0.29	0.10	N	N	N	Y	N	N	Y	Y
Campbell & Mayer, 2009, Study 1	43	43	0.38	0.30	N	Y	Y	N	N	N	Y	Y
		43	0.32	0.30	N	Y	Y	N	N	N	Y	Y
		43	-0.27	0.30	N	Y	Y	N	Y	N	Y	Y
		43	1.20	0.32	N	Y	Y	N	Y	N	Y	Y
Campbell & Mayer, 2009, Study 2	38	38	0.03	0.31	N	Y	Y	N	N	N	Y	Y
		38	0.27	0.32	N	Y	Y	N	Y	N	Y	Y
		38	0.72	0.33	N	Y	Y	N	N	N	Y	Y
		38	0.07	0.31	N	Y	Y	N	Y	N	Y	Y
Christopherson, 2011	40	40	-0.25	0.31	N	Y	N	Y	N/A	N	Y	Y
		40	0.86	0.32	N	Y	N	Y	N/A	N	Y	Y
		40	0.54	0.31	N	Y	N	Y	N/A	N	Y	Y
		40	-0.27	0.31	N	Y	N	Y	N/A	N	Y	Y
		40	1.15	0.33	N	Y	N	Y	N/A	N	Y	Y
Deslauriers, Schelew, & Wieman, 2011	382	382	2.31	0.13	N	Y	N	Y	N	Y	Y	Y
Doucet, Vrins, & Harvey, 2009	169	169	0.03	0.15	Y	Y	N	N	Y	N	Y	Y
		169	0.42	0.16	N	Y	N	Y	N/A	N	Y	Y
Elashvili, Denehy, Dawson, & Cunningham, 2008	76	76	0.42	0.23	Y	Y	Y	N	N	N	Y	Y
		76	-0.33	0.23	N	Y	Y	N	N	N	Y	Y
		76	0.29	0.23	Y	Y	Y	N	N	N	Y	Y
		74	0.02	0.23	Y	Y	Y	N	N	N	Y	Y
		74	1.04	0.25	N	Y	Y	N	N	N	Y	Y
		74	0.14	0.23	Y	Y	Y	N	N	N	Y	Y
FitzPatrick, Finn, & Campisi, 2011, Study 2	151	151	0.43	0.17	N	N	N	N/A	N/A	Y	Y	Y
FitzPatrick et al., 2011, Study 3	115	115	0.36	0.19	N	N	N	N/A	N/A	Y	Y	Y
FitzPatrick et al., 2011, Study 4	59	59	-0.85	0.27	N	N	N	N/A	N/A	Y	Y	Y
		59	-1.19	0.28	N	N	N	N/A	N/A	Y	Y	Y
Geburu, Phelps, & Wulfsberg, 2012	92	92	0.18	0.21	Y	N	N	N/A	N	N	Y	Y
Gray & Steer, 2012	126	126	0.05	0.18	N	Y	N	Y	N	Y	Y	Y
		126	1.04	0.20	N	Y	N	Y	N	Y	Y	Y
		126	0.06	0.18	N	Y	N	Y	N	Y	Y	Y
Knapp & Desrochers, 2009	41	36	-0.07	0.33	Y	Y	Y	Y	N	N	Y	Y
		41	0.66	0.31	N	Y	Y	Y	N	N	Y	Y
Lim, 2011	56	56	0.29	0.27	N	Y	Y	Y	N	Y	Y	Y
		56	0.24	0.27	Y	Y	Y	Y	N	Y	Y	Y
Lin, Liu, & Chu, 2011	275	275	0.20	0.16	N	Y	N	N/A	N	N	Y	Y
		275	0.52	0.16	N	Y	N	N/A	N	N	Y	Y
Liu, Gettig, & Fjortoft, 2010	179	179	0.00	0.15	Y	Y	Y	Y	N	N	Y	Y
		179	0.34	0.15	N	Y	Y	Y	N	N	Y	Y
Martyn, 2007	92	92	-0.17	0.21	N	Y	N	N/A	N/A	N	Y	Y
Mayer et al., 2009	250	250	0.26	0.13	N	Y	N	N	Y	N	Y	Y
		246	0.18	0.13	N	Y	N	Y	N	N	Y	Y
		250	0.38	0.13	N	Y	N	N	Y	N	Y	Y
		246	0.45	0.13	N	Y	N	Y	N	N	Y	Y
McCurry & Revell, 2011	64	64	0.80	0.26	N	N	N	N/A	N	N	Y	Y
Miller, Ashar, & Getz, 2003	283	283	-0.31	0.12	N	Y	Y	Y	N	N	Y	Y
Patterson, Kilpatrick, & Woebkenberg, 2010	70	70	-0.14	0.24	N	Y	N	Y	N/A	N	Y	Y
		70	-0.24	0.24	N	Y	N	Y	N/A	N	Y	Y
		70	0.45	0.24	N	Y	N	Y	N/A	N	Y	Y
		70	0.23	0.24	N	Y	N	Y	N/A	N	Y	Y
Plant, 2007	36	36	0.21	0.32	N	Y	N	N	N	N	Y	Y
		14	-0.05	0.51	Y	Y	N	N	N	N	Y	Y
Pradhan, Sparano, & Ananth, 2005	17	17	0.61	0.48	Y	Y	N/A	N/A	N/A	N	Y	Y
Radosevich, Salomon, Radosevich, & Kahn, 2008	145	145	0.90	0.17	Y	Y	N	N	N/A	N	Y	Y
		145	0.40	0.17	N	Y	N	N	N/A	N	Y	Y
Rubio, Bassignani, White, & Brant, 2008	22	19	1.77	0.53	Y	Y	Y	N	N	N	Y	Y
		22	0.88	0.43	N	Y	Y	N	N	N	Y	Y
Shaffer & Collura, 2009	92	92	0.45	0.21	N	N	Y	Y	Y	N	Y	Y
Tregonning, Doherty, Hornbuckle, & Dickinson, 2012	126	103	0.44	0.20	Y	N	N	N/A	N	N	Y	Y
		126	0.56	0.18	N	N	N	N/A	N	N	Y	Y
		111	-0.26	0.19	Y	N	N	N/A	N	N	Y	Y

Table 3 (continued)

Study	n_{total}	$n_{pairwise}$	g_{raw}	SE_{raw}	Delayed testing	Baseline control	One-shot	Equivalent exposure	Repeated questions	Peer discussion	Display	Elaboration
	115	0.53	0.19	N	N	N	N/A	N	N	Y	Y	
Yourstone, Kraye, & Albaum, 2008, Instructor A	98	98	0.42	0.20	N	Y	N	Y	N	N	Y	Y
Yourstone et al., 2008, Instructor B	92	92	0.39	0.20	N	Y	N	Y	N	N	Y	Y
	92	92	0.33	0.21	N	Y	N	Y	N	N	Y	Y
		92	0.55	0.21	N	Y	N	Y	N	N	Y	Y

Note. N/A, not available.

Table 4

Descriptive statistics of study features.

Study features	Yes (%)	No (%)	N/A (%)	Total
Published after 2007	67 (93)	5 (7)	0 (0)	72
Using undergraduates/graduates as sample	67 (93)	5 (7)	0 (0)	72
Using STEM-related learning materials	42 (58)	30 (42)	0 (0)	72
Controlling for baseline	56 (78)	16 (22)	0 (0)	72
Administering delayed post-tests	15 (21)	57 (79)	0 (0)	72
One-shot intervention	25 (35)	46 (64)	1 (1)	72
Administering in-class questions to the control group	32 (44)	24 (33)	16 (22)	72
Repeating in-class questions as post-tests	8 (11)	46 (64)	18 (25)	72
Allowing subjects to discuss with peers	11 (15)	61 (85)	0 (0)	72
Displaying class results	72 (100)	0 (0)	0 (0)	72
Elaborating correct/incorrect answers	71 (99)	1 (1)	0 (0)	72

Note. N/A, not available.

between groups. The current meta-analysis hence set the baseline-control variable as a moderator to explore the relation between the outcomes measured and the research defect. As shown in Table 5, a total of 30 combined effect sizes were obtained from the comparisons using immediate post-tests, and 11 combined effect sizes were of delayed post-tests. On immediate post-tests, the studies controlling for baseline produced a mean effect size favoring clicker-integrated instruction ($g = 0.49$, 95% CI [0.18, 0.80], $z = 3.10$, $p < .01$), which was approximately twice as large as that of the studies without baseline control ($g = 0.24$, 95% CI [-0.08, 0.57], $z = 1.46$, $p = .14$). This trend also emerged from the comparisons using delayed post-tests. Though the mean effect size of baseline-controlled studies decayed, it remained significant ($g = 0.34$, 95% CI [0.02, 0.66], $z = 2.06$, $p = .04$). The mean effect size of the non-control studies was rather small and insignificant ($g = 0.13$, 95% CI [-0.15, 0.41], $z = 0.95$, $p = .34$), but this result was not informative because the sample size was too small ($n = 2$). Considering such the significant relationship between the research defect and outcomes, the studies without baseline control were excluded from the following analysis. Overall, the general outcomes accompanied with clicker-integrated instruction were greater than those of conventional lectures, regardless what specific instructional strategies were used in clicker-integrated instruction or whether the outcomes were assessed by immediate or delayed post-tests. The mean effect sizes were statistically and practically significant.

Publication bias should be taken into account in any meta-analysis, given that studies with statistically significant positive results are more likely to get published by academic journals than those with negative or statistically non-significant results. Multiple methods were used to examine whether the mean effect size of the selected studies was significantly overestimated. The result of Egger's regression test (Egger, Smith, Schneider, & Minder, 1997) indicated that the asymmetry in the distribution of effect sizes was not statistically significant ($t = 0.16$, $p = .88$). Duval and Tweedie's trim-and-fill method (2000) also indicated that no studies, which had effect sizes smaller than the mean effect, was needed to add into the meta-analysis. Furthermore, the overall analysis of this review yielded a classic fail-safe N of 966, meaning that to make the mean effect size become insignificant, 966 studies with null results would be needed. Given that we were able to identify only 72 pairwise comparisons that looked at the relative learning gains between clicker-integrated instruction and conventional lectures, it is unlikely that nearly 1000 studies were missing. Therefore, it is unlikely that the mean effects sizes were significantly overestimated.

Table 5

Mean effect sizes by baseline control and delayed testing.

Baseline control	Delayed testing									
	No					Yes				
	n	M	SE	95% CI		n	M	SE	95% CI	
			Lower	Upper				Lower	Upper	
No	8	0.24	0.17	-0.08	0.57	2	0.13	0.14	-0.15	0.41
Yes	22	0.49	0.16	0.18	0.80	9	0.34	0.16	0.02	0.66

It should be noted that the overall effect size used two types of lectures as the control group. The first one was the lecture without any in-class questions. The other one was the lecture with in-class questions. Some researchers may consider the comparison as uneven if the control group received no in-class questions, which we termed as *unequal-exposure to test items* in Section 2.2. To deal with this issue, we compared the mean effect sizes obtained from the two types of control groups (i.e., lecturing with vs. without in-class questions). The results indicated that unequal-exposure to in-class questions between groups was not a crucial factor to determine the outcomes of primary studies. Please refer to Section 4.4 for more details.

4.3. Relationship between instruction length and learning outcomes

As shown in Table 6, across the primary studies in this field, the relative learning gains of clicker-integrated instruction did not decrease with the extension of treatment duration. Rather, if clicker-integrated instruction was conducted under relatively long durations, the mean effect size of relative learning gains increased from 0.33 (95% CI [0.05, 0.62], $z = 2.27$, $p = .02$) to 0.57 (95% CI [0.12, 1.01], $z = 2.52$, $p = .01$) on immediate post-tests. A similar trend also emerged from the results on delayed post-tests, but it was not statistically significant and the total sample size was too small ($g = 0.35$, 95% CI [-0.36, 1.06], $z = 0.97$, $p = .33$, $n = 3$ for the studies lasting longer than one lecture; $g = 0.27$, 95% CI [-0.13, 0.66], $z = 1.33$, $p = .18$, $n = 5$ for the one-lecture studies). Moreover, as shown in Table 7, such a boost in learning gains was prominent when outcomes were assessed immediately by a set of new questions, which was relevant to, but not identical to, in-class questions ($g = 0.75$, 95% CI [0.20, 1.29], $z = 2.66$, $p < .01$).

4.4. Relationship between question-answering activities and learning outcomes

As mentioned in Section 2.2, indeed, some comparative studies administrated in-class questions only to the experimental group, simply repeated the questions as a post-test, and merely used students' mean scores on the post-test to argue the usefulness of clickers. Although the contribution of this kind of studies to the theory and practice of clicker-integrated instruction would be rather limited, they certainly provided with the present meta-analysis an empirical baseline to infer whether clicker-integrated instruction went beyond rote memorization of in-class questions. As shown in Table 8, on immediate post-tests, clicker-integrated instruction yielded a greater learn gain than lectures. A medium effect size was obtained when the control groups received lectures without in-class questions ($g = 0.55$, 95% CI [0.23, 0.87], $z = 3.41$, $p < .001$). The effect size remained approximately the same ($g = 0.49$, 95% CI [0.04, 0.94], $z = 2.13$, $p = .03$) if the control groups received in-class questions during lectures. Such results signaled that unequal-exposure to post-test items might not be a crucial factor to determine the outcomes of primary studies. Since the total sample size of delayed post-tests was rather small, we tended to focus on the results of immediate post-tests.

Not surprisingly, if immediate post-tests simply repeated the clicker questions, clicker-integrated instruction produced positive outcomes in the unequal-exposure condition. As shown in Table 9, the mean effect size was significant ($g = 0.28$, 95% CI [0.06, 0.50], $z = 2.44$, $p = .02$). However, there were two findings, which deserve more attention. When the test items which were exactly the same as in-class questions were removed from post-tests, the mean effect size was still significant ($g = 0.60$, 95% CI [0.22, 0.98], $z = 3.09$, $p < .01$), and was twice as large as the effect size obtained from repeated post-tests. Furthermore, the mean effect size almost remained the same even when the control group received in-class questions as well ($g = 0.56$, 95% CI [0.00, 1.12], $z = 1.95$, $p = .05$). In other words, the enhancement of learning gains from clicker-integrated instruction was not likely a result of unequal exposure to the content of clicker questions between experimental and control groups. Nonetheless, it should be noted that the aforementioned claims are valid only for the learning outcomes assessed by immediate post-tests; the amount of studies using new questions to construct delayed post-tests is too little to make reliable inferences.

4.5. Relationship between peer discussion and learning outcomes

Another factor which significantly correlated with the positive outcome of clicker-integrated instruction was asking subjects to generate explanations and justifications for their own answers to clicker questions. As shown in Table 10, clicker-integrated instruction in conjunction with peer discussion, which encouraged subjects to articulate their thought processes to

Table 6
Mean effect sizes for baseline-controlled studies by one-shot and delayed testing.

One-shot	Delayed testing									
	No					Yes				
	n	M	SE	95% CI		n	M	SE	95% CI	
				Lower	Upper				Lower	Upper
Yes	9	0.33	0.15	0.05	0.62	5 ^a	0.27	0.20	-0.13	0.66
No	13	0.57	0.22	0.12	1.01	3 ^a	0.35	0.36	-0.36	1.06

Note. ^a The number of baseline-controlled studies using delayed post-tests decreases from 9 to 8 because the study of Pradhan et al. (2005) is excluded; they did not specify the length or duration of their intervention.

Table 7

Mean effect sizes for baseline-controlled studies using immediate post-tests by one-shot and repeated questions.

One-shot	Repeated questions									
	No					Yes				
	<i>n</i>	<i>M</i>	<i>SE</i>	95% CI		<i>n</i>	<i>M</i>	<i>SE</i>	95% CI	
				Lower	Upper				Lower	Upper
Yes	9 ^a	0.34	0.15	0.05	0.62	2 ^a	0.32	0.23	-0.12	0.77
No	9 ^a	0.75	0.28	0.20	1.29	1 ^a	0.22	0.13	-0.04	0.48

Note. ^a Studies of Christopherson (2011), Martyn (2007), Patterson et al. (2010), and Radosevich et al. (2008) are excluded from this analysis because they did not specify the content of post-tests. Both of Campbell and Mayer's Study 1 and Study 2 (2009) and Mayer et al.'s study (2009) are split into 2 parts. Their post-tests consisted of two sections; section one simply repeated the in-class clicker questions, whereas section two was a set of new questions. Therefore, the amount of baseline-controlled studies using immediate post-tests decreases from 21 to 20 (22-4 + 3 = 21).

Table 8

Mean effect sizes for baseline-controlled studies by equivalent exposure and delayed testing.

Equivalent exposure	Delayed testing									
	No					Yes				
	<i>n</i>	<i>M</i>	<i>SE</i>	95% CI		<i>n</i>	<i>M</i>	<i>SE</i>	95% CI	
				Lower	Upper				Lower	Upper
No	8 ^a	0.55	0.16	0.23	0.87	5	0.52	0.27	-0.02	1.05
Yes	13 ^a	0.49	0.23	0.04	0.94	3	0.04	0.12	-0.20	0.28
n/a	2 ^a	0.11	0.26	-0.41	0.63	1	0.64	0.50	-0.34	1.62

Note. ^a The number of baseline-controlled studies using immediate post-tests increases from 22 to 23 because Mayer et al.'s study (2009) is split into 2 parts; their study compared the clicker group with two different groups, including (1) the control group with paper-based in-class questions and (2) the control group without in-class questions.

Table 9

Mean effect sizes for baseline-controlled studies using immediate post-tests by equivalent exposure and repeated questions.

Equivalent exposure	Repeated questions									
	No					Yes				
	<i>n</i>	<i>M</i>	<i>SE</i>	95% CI		<i>n</i>	<i>M</i>	<i>SE</i>	95% CI	
				Lower	Upper				Lower	Upper
No	7	0.60	0.19	0.22	0.98	3	0.28	0.11	0.06	0.50
Yes	10	0.56	0.29	0.00	1.12	1	0.18	0.13	-0.07	0.43

Note. Studies of Christopherson (2011), Martyn (2007), Patterson et al. (2010), and Radosevich et al. (2008) are excluded from this analysis because they did not specify the content of post-tests. Mayer et al.'s study (2009) is further split into 2 parts because their post-tests consisted of two sections; section one simply repeated the in-class clicker questions, whereas section two was a set of new questions. Study 1 and 2 of Campbell and Mayer (2009) are split into 2 parts as well because they used the same design of post-test measures.

Table 10

Mean effect sizes for baseline-controlled studies using immediate post-tests by peer discussion and repeated questions.

Peer discussion	Repeated questions									
	No					Yes				
	<i>n</i>	<i>M</i>	<i>SE</i>	95% CI		<i>n</i>	<i>M</i>	<i>SE</i>	95% CI	
				Lower	Upper				Lower	Upper
No	13 ^a	0.34	0.10	0.15	0.53	3 ^a	0.25	0.11	0.02	0.47
Yes	4 ^a	1.19	0.58	0.06	2.32	0 ^a	None	None	None	None

Note. ^a The number of baseline-controlled studies using immediate post-tests decreases from 22 to 20 because the studies without specifying the content of post-tests are excluded from this analysis.

peers immediately after responding to clicker questions, produced a strong positive outcome on immediate post-tests. The mean effect size was statistically significant with a pretty large magnitude ($g = 1.19$, 95% CI [0.06, 2.32], $z = 2.04$, $p = .04$). Clicker-integrated instruction without peer discussion yielded a positive mean effect size as well, but the magnitude was much smaller than those of peer discussion ($g = 0.34$, 95% CI [0.15, 0.53], $z = 3.53$, $p < .001$). It should be noted these results were derived from the post-tests which did not repeat clicker questions. None of the studies of peer discussion duplicated in-

class questions as post-tests. The result on delayed post-tests was absent because only one study allowed subjects to discuss with peers.

Acknowledging that the mean effect size of clicker-integrated instruction decreased dramatically if the studies involving peer discussion were excluded, we wondered if the superior learning gains of clicker-integrated instruction, in comparison with the conventional lectures with in-class questions, was primarily due to the generation of explanations or justifications. The large effect sizes reported by the studies of peer discussion may bias the mean effect size which used conventional lectures with in-class questions as the control group. This was probably the reason why the mean effect size, using conventional lectures with in-class questions as the control group, was similar to the mean effect size using conventional lectures without in-class questions as the control group. However, as indicated by a follow-up analysis, even if the studies involving peer discussion were excluded, the mean effect size comparing to the lectures with in-class questions ($g = 0.33$, 95% CI [0.04, 0.63], $z = 2.21$, $p = .03$) was as large as the effect size comparing to those without in-class questions ($g = 0.39$, 95% CI [0.20, 0.57], $z = 4.04$, $p < .001$). The difference between these two effect sizes was negligibly small ($g_{\text{diff}} = 0.06$), providing strong evidence that the superiority of clicker-integrated instruction over conventional lectures did not result from unequal exposure to in-class questions.

5. Discussion

The meta-analysis results suggest that the superior effect of clicker-integrated instruction, compared to conventional lectures, stands on firm empirical ground. As pointed out by Lantz (2010) and Lasry (2008), a major criticism to clicker adoption is that clickers might be nothing but a mere novelty for students. Undoubtedly, the effectiveness of educational technology will always suffer from the novelty effect because students often get excited when new technology is instituted (Bangert-Drowns, Kulik, & Kulik, 1985; Cheung & Slavin, 2013; Kulik & Kulik, 1991; Kulik et al., 1985). In this case, it is more realistic and practical to examine *whether the improvement in learning diminishes significantly as students become more familiar with clickers*, rather than whether there exists a novelty effect across past studies. The meta-analysis results have demonstrated that the superior outcomes of clicker-integrated instruction remained robust even if students had experienced clicker usage for a relatively long duration. Therefore, the novelty effect should not be a primary concern regarding clicker adoption.

The meta-analysis results are also helpful to settle the *clicker vs. method* debate. A sharp schism, between clickers and instructional methods, is emphasized in previous studies (e.g., Anthis, 2011; Campbell & Mayer, 2009; Christopherson, 2011; Mayer et al., 2009; Shapiro & Gordon, 2012). It is argued that the effectiveness of clicker-integrated instruction is falsely attributed to clickers. Rather, the true effect is due to the addition of question-answering activities (i.e., the method) that primes testing and adjunct-question effects (Anthis, 2011; Campbell & Mayer, 2009; Christopherson, 2011; Mayer et al., 2009). It is true that some comparative studies were confounded with question-answering activities because they did not require the control group to respond to questions during class. However, the synthesis of previous studies still yielded a significant positive mean effect size when clicker-integrated instruction was compared with the lectures which also gave students question-answering activities. The mean effect size even increased if learning outcomes were assessed by a set of questions that were different from clicker questions. These results strongly suggest that the use of clickers indeed has some advantages over conventional lectures. Both of testing and adjunct-question effects are inadequate to explain the superiority of clicker-integrated instruction. Moreover, the effectiveness of clicker-integrated instruction appears to go beyond rote memorization of in-class questions; it facilitates knowledge application.

Based on the meta-analysis results, the feedback-intervention effect seems to be more suitable than testing/adjunct-question effects to account for the superiority of clicker-integrated instruction. It is noted that all selected studies did provide students with instant feedback after students responded to in-class questions. As documented in previous literature (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008), feedback on learning performance can be ineffective, or even detrimental, if it threatens students' self-esteem. Compared with the feedback given in conventional lectures, the feedback given in clicker-integrated instruction is less threatening to students' self-esteem because their performance is evaluated in a private way. The feedback given in clicker-integrated instruction is thus more likely to direct students' attention to the focal task and, consequently, improve learning. Previous reviews (e.g., Boscardin & Penuel, 2012; Caldwell, 2007; Fies & Marshall, 2006; Kay & LeSage, 2009) have also indicated that students highly appreciate the anonymity of clickers that enables them to gather feedback without being publicly judged by others. It will be more fruitful for future studies if the research question is reshaped as *how a specific instructional method might be enhanced by the use of clickers*, rather than questioning whether the effectiveness of clicker-integrated instruction should be solely attributed to the instructional method or the clicker.

Some researchers may argue that students' awareness of their responses being collected by clickers significantly contributes to the superior outcomes of clicker-integrated instruction. All kinds of clicker responses, even null responses, are recorded and tracked by the teacher. Students thus increase their attention and consequently, reinforce testing/adjunct-question effects. However, there is a premise required to make this conjecture valid: students must feel that clicking no answer or wrong answers is detrimental for the teacher's evaluation of their performance. Therefore, we tend to further this conjecture as the grading-incentive issue. Most studies included in this review (89%) did not count students' clicker responses as part of a final grade, nor did they explicitly tell students that clicker responses would be graded. On the contrary, some studies even explicitly told students that all responses were anonymous and would not be graded. The study of James (2006) further indicates that students are more likely to mindlessly pick the most popular answers, rather than being cognitively engaged in solving questions, when they are aware that their clicker responses are being graded. Therefore, we contend that

there is no evidence to support that students' awareness of clicker data collection inevitably improves their attention and consequently learning outcomes.

Implementing peer discussion of clicker questions is highly recommended because it is generally able to produce large learning gains. Such a practice has been advocated as the standard model to use clickers in the science classroom (Beatty, Gerace, Leonard, & Dufresne, 2006; Caldwell, 2007; Crouch & Mazur, 2001; Newbury & Heiner, 2012; Wieman et al., 2009). However, it should be noted that peer discussion can be implemented totally without clickers. Therefore, this review does not intend to conclude that clickers facilitate peer discussion. Rather, peer discussion, which primes students to explain their reasoning, enhances the effectiveness of clicker-integrated instruction. The theory of scaffolding (Bruner, 1985; Vygotsky, 1978) might be another candidate to explain the effectiveness of peer discussion, in addition to the self-explanation effect. Through interacting with a more knowledgeable other, a student is enabled to achieve learning outcomes that are beyond his/her independent efforts. The student then internalizes the learning outcomes as the support from the more knowledgeable other is gradually withdrawn. However, the information available in the selected studies is too vague to identify the core components of the scaffolding theory, such as the presence of a more knowledgeable other and how the more knowledgeable other participates in students' learning. Future studies are needed to investigate how students interact with peers within the context of clicker-integrated instruction. Research of this line will also be helpful to understand how the use of clickers may mediate the process and outcomes of peer discussion.

6. Conclusion

The results of meta-analysis suggest that clickers have pedagogical value beyond the novelty effect and simple memorization of in-class questions. Compared with testing and adjunct-question effects, the feedback-intervention effect is a more reasonable theoretical explanation for the superiority of clicker-integrated instruction. Moreover, engaging students in peer discussion is found to be a very promising strategy to promote the effectiveness of clicker-integrated instruction.

Based on the results and discussion, we recommend educators and researchers view learning as a self-regulatory process (Butler & Winne, 1995; Pintrich & Zusho, 2002; Zimmerman, 2001). Students proactively seek, produce, and interpret information as feedback on their learning performance. Feedback changes students' locus of attention and therefore their cognition, motivation, and behavior during learning. Effective instruction should assist students in monitoring and regulating themselves to close the gap between actual and desired performance (Butler & Winne, 1995; Pintrich & Zusho, 2002; Zimmerman, 2001). Such a perspective of learning accommodates the instructional methods and theoretical phenomena reviewed in this article. For instance, the question-answering activity is implemented to help students gauge their progress and gather feedback at the micro level, and in a timely fashion. The implementation of question-answering activity also primes testing and adjunct-question effects that enhance students' memory and direct their attention to instructional material. Peer discussion enables students to share the monitoring and regulatory processes with classmates. It also primes self-explanation effect that prompts students to detect and correct errors in their reasoning.

The perspective of learning as a self-regulatory process also helps educators and researchers to envision how clickers can enhance instructional methods. The central issue is how to use clickers to craft and deliver information that students may find constructive to monitor and regulate cognition, motivation, and behavior during learning. As discussed previously, the anonymous feature of clickers prevents students' weakness from being publicly exposed, while they are gathering information to gauge their own learning. The teacher's feedback following the question is thus more likely to be adapted by students to keep persistency in learning and improving performance. The real-time display of clicker voting results, which is shown before the correct answer is highlighted by the teacher, is also a source of information for students to monitor their performance with reference to peers. Our preliminary study (Chien, Lee, Li, & Chang, 2015) has demonstrated that displaying the voting results can guide students' discussion processes and thus influence learning outcomes. However, the display of clicker voting results may have both positive and negative impacts on students' self-regulatory processes. On the positive side, the display of clicker voting results might be perceived as an "I am not the only one confused" comfort to students (Hoekstra, 2008; Hoekstra & Mollborn, 2012). It helps eliminate the feeling of hopelessness and keep students in regulating themselves to approach the learning goal. The display of clicker voting results may also fuse students to discuss on why some answers are commonly chosen, and thus further engage students in evaluating and correcting each other's reasoning (Hoekstra, 2008; Hoekstra & Mollborn, 2012). On the negative side, students may passively conform to the majority's opinion represented in the clicker voting results, rather than actively examining the flaws of their own reasoning (Nielsen, Hansen-Nygård, & Stav, 2012; Perez et al., 2010). Further studies are needed to clarify what kind of voting results (e.g., univocal idea or diverse ideas represented in the voting results) can facilitate, or even inhibit, productive self-regulation.

The use of clicker-like teaching strategies has practical and critical relevance to 21st century learning systems such as Massive Open Online Courses (MOOCs). In MOOCs, students are asked to watch pre-shot video clips to learn course content by themselves. Periodically prompting students to answer the questions related to the course content, while they are watching the video clips, may prime students to keep, monitor and regulate their own learning. Since MOOCs are built on the online platforms, it is easy to provide students with an anonymous mechanism to respond to questions and receive feedback. The social media embedded in MOOCs is also capable of offering students an anonymous mechanism to participate in peer discussion, either in synchronous or asynchronous forms.

It should be also noted that almost all of the selected studies used undergraduates or graduates as research samples. Although such the homogeneity in subjects is logically good for estimating effect sizes more precisely, the generalizability of

results at the same time is limited to the instruction conducted with undergraduates or graduates. Studies with high-school or elementary students are recommended to verify the generalizability of the findings. Furthermore, the conclusion made in this review should be limited to the learning outcomes assessed by immediate post-tests because the amount of studies using delayed post-tests is too little to make reliable inferences. Studies assessing learning gains, at various time points, are recommended to examine the stability of the positive learning outcomes of clicker-integrated instruction. Reporting on content of in-class questions, post-tests, and teachers' feedback is highly recommended for future research practice. The proposed information gathered will be helpful in gaining a deeper understanding of how to design appropriate clicker questions and feedback to facilitate learning outcomes.

Acknowledgements

This research is supported by the Aim for the Top University Project of National Taiwan Normal University (sponsored by the Ministry of Education, Taiwan, R.O.C.), the International Research-Intensive Center of Excellence Program (sponsored by National Taiwan Normal University and the Ministry of Science and Technology, Taiwan, R.O.C., under grant no. MOST 104-2911-I-003-301), and the Ministry of Science and Technology, Taiwan, R.O.C. (under grant no. MOST 102-2511-S003-052-MY3). The authors gratefully acknowledge the assistance of Terrence Wong in editing the manuscript.

References¹

- *Agbatogun, A. O. (2012). Exploring the efficacy of student response system in a sub-saharan african country: a sociocultural perspective. *Journal of Information Technology Education: Research*, 11, 249–267.
- Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26(2), 147–179.
- Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research*, 49(2), 280–318.
- Anthis, K. (2011). Is it the clicker, or is it the question? untangling the effects of student response system use. *Teaching of Psychology*, 38(3), 189–193.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, 95(4), 774–783.
- Chien, Y. T., Lee, Y. H., Li, T. Y., & Chang, C. Y. (2015). Examining the effects of displaying clicker voting results on high school students' voting behaviors, discussion processes, and learning outcomes. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(5), 1089–1104.
- Azevedo, R., & Aleven, V. A. (2013). *International handbook of metacognition and learning technologies* (Vol. 26). New York, NY: Springer.
- *Bachman, L., & Bachman, C. (2011). A study of classroom response system clickers: increasing student engagement and performance in a large undergraduate lecture class on architectural research. *Journal of Interactive Learning Research*, 22(1), 5–21.
- Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1985). Effectiveness of computer-based education in secondary schools. *Journal of Computer-Based Instruction*, 12(3), 59–68.
- Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61(2), 213–238.
- *Bartsch, R. A., & Murphy, W. (2011). Examining the effects of an electronic classroom response system on student engagement and performance. *Journal of Educational Computing Research*, 44(1), 25–33.
- Beatty, I. D., Gerace, W. J., Leonard, W. J., & Dufresne, R. J. (2006). Designing effective questions for classroom response system teaching. *American Journal of Physics*, 74(1), 31–39.
- Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13(2), 221–252.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. In *Handbook I: Cognitive domain*. New York, NY: David McKay.
- de Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students' academic performance: a meta-analysis. *Review of Educational Research*, 84(4), 509–545.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons.
- Boscardin, C., & Penuel, W. (2012). Exploring benefits of audience-response systems on learning: a review of the literature. *Academic Psychiatry*, 36(5), 401–407.
- Bruner, J. S. (1985). Vygotsky: a historical and conceptual perspective. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 21–34). Cambridge, England: Cambridge University Press.
- *Butler, M., Pyzdrowski, L., Walker, V., & Yoho, S. (2010). Studying personal response systems in a college algebra course. *Investigations in Mathematics Learning*, 2(2), 1–18.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3), 245–281.
- Caldwell, J. E. (2007). Clickers in the large classroom: current research and best-practice tips. *CBE-Life Sciences Education*, 6(1), 9–20.
- *Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: does it affect learning from lectures? *Applied Cognitive Psychology*, 23(6), 747–759.
- Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory*, 18(1), 49–57.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III (2006). Retrieval-induced facilitation: initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135(4), 553–571.
- Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: a meta-analysis. *Educational Research Review*, 9, 88–113.
- Chi, M. T. H., & Bassok, M. (1989). Learning from examples via self-explanations. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 251–282). Hillsdale, NJ: Erlbaum.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, 13(2), 145–182.
- Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477.
- Chi, M. T. H., & VanLehn, K. A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences*, 1(1), 69–105.
- Chi, M., & VanLehn, K. (2010). Meta-cognitive strategy instruction in intelligent tutoring systems: how, when, and why. *Educational Technology & Society*, 13(1), 25–39.

¹ References marked with an asterisk indicate studies included in the meta-analysis.

- *Christopherson, K. M. (2011). Hardware or wetware: what are the possible interactions of pedagogy and technology in the classroom? *Teaching of Psychology*, 38(4), 288–292.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445–459.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research*, 52(2), 291–302.
- Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: ten years of experience and results. *American Journal of Physics*, 69(9), 970–977.
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (2nd ed., pp. 317–344). San Diego, CA: Academic Press.
- *Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332(6031), 862–864.
- Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: a meta-analysis. *Educational Research Review*, 11, 1–26.
- *Doucet, M., Vrins, A., & Harvey, D. (2009). Effect of using an audience response system on learning environment, motivation and long-term retention, during case-discussions in a large group of undergraduate veterinary clinical pharmacology students. *Medical Teacher*, 31(12), E570–E579.
- Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634.
- *Elashvili, A., Denehy, G. E., Dawson, D. V., & Cunningham, M. A. (2008). Evaluation of an audience response system in a preclinical operative dentistry course. *Journal of Dental Education*, 72(11), 1296–1303.
- Fies, C., & Marshall, J. (2006). Classroom response systems: a review of the literature. *Journal of Science Education and Technology*, 15(1), 101–109.
- *FitzPatrick, K. A., Finn, K. E., & Campisi, J. (2011). Effect of personal response systems on student perception and academic performance in courses in a health sciences curriculum. *Advances in Physiology Education*, 35(3), 280–289.
- Frase, L. T. (1967). Learning from prose material: length of passage, knowledge of results, and position of questions. *Journal of Educational Psychology*, 58(5), 266–272.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18.
- *Geburu, M. T., Phelps, A. J., & Wulfsberg, G. (2012). Effect of clickers versus online homework on students' long-term retention of general chemistry course material. *Chemistry Education Research and Practice*, 13(3), 325–329.
- Gilbert, A. (2005). *New for back-to-school: 'Clickers'*. Retrieved October 28, 2015 http://news.cnet.com/New-for-back-to-school-Clickers/2100-1041_3-5819171.html.
- Goldberg, B., & Spain, R. (2014). Creating the intelligent novice: supporting self-regulated learning and metacognition in educational technology. In R. A. Sottolare, A. C. Graesser, X. Hu, & B. S. Goldberg (Eds.), *Instructional management: Vol. 2. Design recommendations for intelligent tutoring systems* (pp. 105–133). Orlando, FL: U.S. Army Research Laboratory.
- Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, 40(4), 225–234.
- *Gray, K., & Steer, D. N. (2012). Personal response systems and learning: it is the pedagogy that matters, not the technology. *Journal of College Science Teaching*, 41(5), 80–88.
- Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, 56(2), 212–242.
- Hamilton, R. J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research*, 55(1), 47–85.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128.
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2), 490–499.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486–504.
- 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.
- Hoekstra, A., & Mollborn, S. (2012). How clicker use facilitates existing pedagogical practices in higher education: data from interdisciplinary research on student response systems. *Learning, Media and Technology*, 37(3), 303–320.
- James, M. C. (2006). The effect of grading incentive on student discourse in peer instruction. *American Journal of Physics*, 74(8), 689–691.
- Järvelä, S., Kirschner, P., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., et al. (2015). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educational Technology Research and Development*, 63(1), 125–142.
- 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.
- Kirk, R. E. (1996). Practical significance: a concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746–759.
- Kirschner, P. A., Kreijns, K., Phielix, C., & Franssen, J. (2015). Awareness of cognitive and social behaviour in a CSCL environment. *Journal of Computer Assisted Learning*, 31(1), 59–77.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- *Knapp, F. A., & Desrochers, M. N. (2009). An experimental evaluation of the instructional effectiveness of a student response system: a comparison with constructed overt responding. *International Journal of Teaching and Learning in Higher Education*, 21(1), 36–46.
- Kreijns, K., Kirschner, P. A., & Vermeulen, M. (2013). Social aspects of CSCL environments: a research framework. *Educational Psychologist*, 48(4), 229–242.
- Kulik, C. L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated analysis. *Computers in Human Behavior*, 7(1–2), 75–94.
- Kulik, J. A., Kulik, C. L. C., & Bangert-Drowns, R. L. (1985). Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior*, 1(1), 59–74.
- Lantz, M. E. (2010). The use of 'Clickers' in the classroom: teaching innovation or merely an amusing novelty? *Computers in Human Behavior*, 26(4), 556–561.
- Lasry, N. (2008). Clickers or flashcards: is there really a difference? *The Physics Teacher*, 46(4), 242–244.
- *Lim, K. H. (2011). Addressing the multiplication makes bigger and division makes smaller misconceptions via prediction and clickers. *International Journal of Mathematical Education in Science and Technology*, 42(8), 1081–1106.
- *Lin, Y. C., Liu, T. C., & Chu, C. C. (2011). Implementing clickers to assist learning in science lectures: the clicker-assisted conceptual change model. *Australasian Journal of Educational Technology*, 27(6), 979–996.
- *Liu, F. C., Gettig, J. P., & Fjortoft, N. (2010). Impact of a student response system on short- and long-term learning in a drug literature evaluation course. *American Journal of Pharmaceutical Education*, 74(1). Article 6.
- MacArthur, J. R., & Jones, L. L. (2008). A review of literature reports of clickers applicable to college chemistry classrooms. *Chemistry Education Research and Practice*, 9(3), 187–195.
- *Martyn, M. (2007). Clickers in the classroom: an active learning approach. *EDUCAUSE Quarterly*, 30(2), 71–74.
- Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: learning from errors with metacognitive tutoring. *Educational Psychologist*, 40(4), 257–265.

- *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.
- Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.
- *McCurry, M. K., & Hunter Revell, S. M. (2011). Evaluating the effectiveness of personal response system technology on millennial student learning. *The Journal of Nursing Education*, 50(8), 471–475.
- *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.
- Nelson, C., Hartling, L., Campbell, S., & Oswald, A. E. (2012). The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME guide no. 21. *Medical Teacher*, 34(6), e386–e405.
- Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? the case of analogical problem-solving. *British Journal of Educational Psychology*, 68(1), 15–24.
- Newbury, P., & Heiner, C. (2012). *Ready, set, react! getting the most out of peer instruction using clickers*. Retrieved October 28, 2015, from http://www.cwsei.ubc.ca/Files/ReadySetReact_3fold.pdf.
- Nielsen, K. L., Hansen-Nygård, G., & Stav, J. B. (2012). Investigating peer instruction: how the initial voting session affects students' experiences of group discussion. *ISRN education*, 2012. Article ID 290157.
- *Patterson, B., Kilpatrick, J., & Woebkenberg, E. (2010). Evidence for teaching practice: the impact of clickers in a large classroom environment. *Nurse Education Today*, 30(7), 603–607.
- Perez, K. E., Strauss, E. A., Downey, N., Galbraith, A., Jeanne, R., & Cooper, S. (2010). Does displaying the class results affect student discussion during peer instruction? *CBE-Life Sciences Education*, 9(2), 133–140.
- Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: effects of peer feedback and reflection. *Computers in Human Behavior*, 26(2), 151–161.
- Phielix, C., Prins, F. J., Kirschner, P. A., Erkens, G., & Jaspers, J. (2011). Group awareness of social and cognitive performance in a CSCL environment: effects of a peer feedback and reflection tool. *Computers in Human Behavior*, 27(3), 1087–1102.
- Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: the role of cognitive and motivational factors. In A. Wigfield, & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 249–284). San Diego, CA: Academic Press.
- *Plant, J. D. (2007). Incorporating an audience response system into veterinary dermatology lectures: effect on student knowledge retention and satisfaction. *Journal of Veterinary Medical Education*, 34(5), 674–677.
- *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(5), 1827–1830.
- *Radosevich, D. J., Salomon, R., Radosevich, D. M., & Kahn, P. (2008). Using student response systems to increase motivation, learning, and knowledge retention. *Innovate: Journal of Online Education*, 5(1). Article 4.
- Raudenbush, S. W. (2009). Analyzing effect sizes: random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295–316). New York: NY: Russell Sage Foundation.
- Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science*, 21(1), 1–29.
- Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: the effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, 23(1), 90–108.
- Rickards, J. P. (1979). Adjunct postquestions in text: a critical review of methods and processes. *Review of Educational Research*, 49(2), 181–196.
- Rittle-Johnson, B. (2006). Promoting transfer: effects of self-explanation and direct instruction. *Child Development*, 77(1), 1–15.
- Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.
- Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.
- *Rubio, E. I., Bassignani, M. J., White, M. A., & Brant, W. E. (2008). Effect of an audience response system on resident learning and retention of lecture material. *American Journal of Roentgenology*, 190(6), W319–W322.
- Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research*, 84(3), 328–364.
- Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect size. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 257–278). New York: NY: Russell Sage Foundation.
- *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.
- Shapiro, A. M., & Gordon, L. T. (2012). A controlled study of clicker-assisted memory enhancement in college classrooms. *Applied Cognitive Psychology*, 26(4), 635–643.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- 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.
- Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: a learning theory perspective. *Educational Research Review*, 9, 1–15.
- *Tregonning, A. M., Doherty, D. A., Hornbuckle, J., & Dickinson, J. E. (2012). The audience response system and knowledge gain: a prospective study. *Medical Teacher*, 34(4), e269–e274.
- VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences*, 2(1), 1–59.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wieman, C., Perkins, K., Gilbert, S., Benay, F., Kennedy, S., Semsar, K., et al. (2009). *Clicker resource guide: An instructor's guide to the effective use of personal response systems (clickers) in teaching*. Vancouver, BC, Canada: University of British Columbia. Available from: http://www.cwsei.ubc.ca/resources/files/Clicker_guide_CWSEL_CU-SEL.pdf.
- Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, 12(2), 233–262.
- *Yourstone, S. A., Kraye, H. S., & Albaum, G. (2008). Classroom questioning with immediate electronic response: do clickers improve learning? *Decision Sciences Journal of Innovative Education*, 6(1), 75–88.
- Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: an overview and analysis. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 1–38). Mahwah, NJ: Lawrence Erlbaum Associates.