Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: informahealthcare.com/journals/imte20

The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME Guide No. 21

Cody Nelson, Lisa Hartling, Sandra Campbell & Anna E. Oswald

To cite this article: Cody Nelson, Lisa Hartling, Sandra Campbell & Anna E. Oswald (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, DOI: 10.3109/0142159X.2012.680938

To link to this article: https://doi.org/10.3109/0142159X.2012.680938

Published online: 11 May 2012.

Submit your article to this journal

Article views: 4836

View related articles

Citing articles: 14 View citing articles

Full Terms & Conditions of access and use can be found at
https://informahealthcare.com/action/journalInformation?journalCode=imte20

2012; 34: e386–e405 MEDICAL TEACHER

WEB PAPER
BEME GUIDE

The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME Guide No. 21

CODY NELSON, LISA HARTLING, SANDRA CAMPBELL & ANNA E. OSWALD
University of Alberta, Canada

Abstract

Background: Audience response systems (ARS) represent one approach to make classroom learning more active. Although ARS may have pedagogical value, their impact is still unclear. This systematic review aims to examine the effect of ARS on learning outcomes in health professions education.

Methods: After a comprehensive literature search, two reviewers completed title screening, full-text review and quality assessment of comparative studies in health professions education. Qualitative synthesis and meta-analysis of immediate and longer term knowledge scores were conducted.

Results: Twenty-one of 1013 titles were included. Most studies evaluated ARS in lectures (20 studies) and in undergraduates (14 studies). Fourteen studies reported statistically significant improvement in knowledge scores with ARS. Meta-analysis showed greater differences with non-randomised study design. Qualitative synthesis showed greater differences with non-interactive teaching comparators and in postgraduates. Six of 21 studies reported student reaction; 5 favoured ARS while 1 had mixed results.

Conclusion: This review provides some evidence to suggest the effectiveness of ARS in improving learning outcomes. These findings are more striking when ARS teaching is compared to non-interactive sessions and when non-randomised study designs are used. This review highlights the importance of having high quality studies with balanced comparators available to those making curricular decisions.

Introduction

There has been a shift in health trainee education from traditional lectures to a more engaging and active style of teaching. This is in part because of the inadequacies of traditional lecturing to meet the needs of growing class sizes; and the increasing evidence that lectures are not effective for solidifying long-term knowledge acquisition or for promoting translation beyond the acquisition of knowledge to its application in both related and different settings (Alexander et al. 2009; Forsetlund et al. 2009). Audience response system(s) (ARS) represent a recent innovation that is being used by an increasing number of educational institutions to facilitate student engagement and learning. It consists of an input device controlled by the learner, a receiver and a display linked to the input that can be controlled by the instructor. ARS were first seen at Cornell and Stanford Universities in the 1960s but were not made available for commercial use until the 1990s. Since that time, this technology has been evolving to meet the needs of the modern classroom (Judson & Sawada 2002; Abrahamson 2006). A more affordable and convenient ARS was marketed in 1999, and in 2003, it started having widespread use in classrooms of higher education (Banks & Bateman 2004; Abrahamson 2006; Kay & LeSage 2009). ARS are being used in a variety of ways: as a learning strategy to facilitate increased attention, interaction, instruction, student preparation and discussion; to motivate students for attendance and participation; and to provide formative and summative knowledge assessments (Kay & LeSage 2009).

Practice points

- ARS may improve knowledge scores and do improve learner reaction.
- Findings are more striking with non-interactive teaching comparators and non-randomised studies.
- In postgraduates, where sleep deprivation is common, ARS may be even more beneficial (further study required).
- This review highlights the importance of having high quality studies with balanced comparators available to those making curricular decisions.

The literature concerning ARS in education has consistently purported that, when used properly, ARS can achieve positive results for participants (Caldwell 2007; Cain & Robinson 2008). However, there has been reluctance in using ARS by many

Correspondence: A.E. Oswald, 562 Heritage Medical Research Centre, University of Alberta, Edmonton, AB T6G 2S2, Canada. Tel: 780-407-8711; Fax: 780-407-6055; E-mail: oswald@ualberta.ca

ISSN 0142–159X print/ISSN 1466–187X online/12/060386–20 © 2012 Informa UK Ltd.
DOI: 10.3109/0142159X.2012.680938

Copyright Informa Healthcare 2012
Not for Sale or Commercial Distribution
Unauthorized use prohibited.
Authorized users can download, display, view and print a single copy for personal use.

teachers and faculties. Some have expressed concerns regarding the time and effort required to prepare new ARS style lectures (Halloran 1995), the cost to faculty and students of implementing the new system and the decreased time available to cover lecture material (Miller et al. 2003; Cain & Robinson 2008).

Although ARS may have real pedagogical value, their impact on learning in health professions education is still unclear. There have been eight reviews published exploring the cost, use and effect of ARS in the broader education literature (Judson & Sawada 2002; Roschelle et al. 2004; Fies & Marshall 2006; Caldwell 2007; Simpson & Oliver 2007; Cain & Robinson 2008; MacArthur & Jones 2008; Kay & Lesage 2009). However, many of these reviews were not systematic and several had inadequate rigour in their methods as discussed below. Many of these reviews address more general populations, including but not exclusively examining health professions education. Some were published nearly a decade ago and are limited by the number of studies they include.

The most recent systematic review by Kay and LeSage examines the different uses of ARS in higher education, includes 52 studies and represents the most thorough and rigorous review to date. The authors reported a number of promising strategies including collecting formative assessment feedback and peer-based instruction. However, of the 52 studies only seven studies related to health professions education, and these studies focussed on teaching strategies to improve the use of ARS rather than on learning outcomes.

Cain and Robinson published a review in 2008 that gave an overview of the current applications of ARS within health trainee education. This was not a systematic review and reported data on only six studies.

Reviews that report learning outcomes have consistently found that learner reaction is positive (Judson & Sawada 2002; Roschelle et al. 2004; Fies & Marshall 2006; Caldwell 2007; Simpson & Oliver 2007; Cain & Robinson 2008; MacArthur & Jones 2008). However, the reviews that reported knowledge outcomes (Judson & Sawada 2002; Fies & Marshall 2006; Caldwell 2007; Cain & Robinson 2008; MacArthur & Jones 2008) reported mixed results, some studies favouring ARS and others not.

Many reviews have highlighted limitations of the current literature. For example, in 2002, Judson and Sawada published a review that concluded the positive effects of ARS on knowledge scores and learner reaction point more to the teaching practices of the instructor than the incorporation of the ARS technology. The review by Fies and Marshall examined the different uses of ARS in education and concluded that much of the current literature compares ARS versus non-ARS teaching sessions that are unequal. They call for research that rigorously assesses ARS with more balanced comparators in a variety of educational settings.

Until this time, there has been a shortage of literature that would allow a high quality methodological review to be performed that focused on health professions education. We however, in the past few years, a substantial number of new articles with this focus have been published. It is now possible to more rigorously assess the effect of ARS on learning in health professions trainees and provide a better understanding of their use in this distinct context.

Methods

Research question

The overall research question for this systematic review is: what are the effects of ARS on learning outcomes in health professions education? This review includes undergraduate and graduate students, clinical trainees and practicing professionals. The effectiveness of educational strategies was measured in terms of the classic Kirkpatrick model (Kirkpatrick & Kirkpatrick 2006) including change in patients' health, change in learners' behaviour, change in learners' skills, change in learners' knowledge, change in learners' attitudes/perceptions and change in learners' reactions. Although it is not explicit in Kirkpatrick's framework, we included learners' self-confidence under the category of learners' attitudes/perceptions.

Search strategy

A comprehensive search strategy was developed by a health science librarian in consultation with the other co-authors. We identified relevant studies from the online databases listed in Table 1 and from other relevant sources as described below.

Two search strategies were used depending on whether the database in question was health related or not. This was done to ensure the inclusion of all relevant studies. The specific terms and search strategies can be found for health-related databases in Table 2 and general databases in Table 3. In addition, the reference lists of all included studies were hand searched, as were those of relevant reviews that were identified during the title screening procedure described below. We also hand-searched the conference proceedings for the Association of American Medical Colleges, the Association of Medical Education in Europe and the Canadian Conference of Medical Education from 2007 to 2009. A separate cited reference search was conducted using Web of Science and SCOPUS for each included study to identify papers where it had been cited. The primary authors of all included studies were contacted by email to determine if they knew of any unpublished, recently published or ongoing studies relevant to the review. The contact information used was extracted from the included papers or from the university directories associated with the primary authors.

Screening and selection of studies

The titles and abstracts generated from the electronic database searches were collated in a Refworks reference management database. They were then screened by two reviewers (AO and CN) to exclude those that obviously did not meet the inclusion criteria or address the question under study. The full texts of the remaining studies were retrieved and a pre-approved inclusion form was applied to each to identify relevant studies. This was done independently by two reviewers (AO and CN), and any disagreements that arose were resolved through discussion, or with the aid of a third reviewer (LH) as required.

Table 1. Included online databases.

Health-related databases	General databases
Medline (1950 to present)	Physical Education Abstracts
EMBASE (1980 to present)	SCOPUS (1823 to present)
PubMed (1950 to present)	Web of Science (1956 to present)
CINAHL (1937to present)	ERIC (1966 to present)
Cochrane Library (various dates to present)	OpenSigle (various years to present)
	Proquest Dissertations and Theses (content dates vary to present)

Note: Databases – note that all searches were limited to 1970 to July 2010.

Table 2. Search terms and strategy.

Health-related databases	
Health trainee education methods	and 'ARS'
exp Education/	'audience response system*'.mp.
or exp Educational Technology/	or 'classroom response system*'.mp.
or 'teaching method*'.mp.	or 'wireless response system*'.mp.
or curriculum.mp.	or 'electronic voting system*'
or 'instructional method*'.mp.	or 'group response system*'
	or 'personal response system*'
	or clicker*
	or iclicker*
	or 'interactive voting system*'
	or 'student response system*'

Note: Limits: English language, human, 1970 to present.

Table 3. Search terms and strategy.

General databases	
'ARS'	and 'Health professions'
'audience response system*'.mp.	medic*
or 'classroom response system*'.mp.	or nurs*
or 'wireless response system*'.mp.	or 'physical therap*'
or 'electronic voting system*'	or physician*
or 'group response system*'	or health
or 'personal response system*'	or dentist*
or clicker*	or pharmac*
or iclicker*	or 'occupational therap*'
or 'interactive voting system*'	or doctor*
or 'student response system*'	or dietitician*
	or psychologist*
	or clinic*

Note: Limits: English language, human, 1970 to present.

The inclusion criteria are detailed in Table 4. These were applied to each potentially relevant study to evaluate whether the study should be included in the review. This review focused on health professions trainees who experienced teaching interventions as evaluated by controlled studies.

Assessment of methodological quality

The methodological quality of included studies was evaluated independently by two reviewers (LH and CN) using well-recognised tools. The Cochrane Risk of Bias tool was used for controlled trials (Higgins & Green 2006). The Newcastle-Ottawa Scale was used for cohort studies (Wells et al.). Discrepancies were resolved through consensus.

Data extraction

Data were extracted and entered into an electronic data extraction form. These were developed and piloted in a systematic review performed by the authors (Hartling et al. 2010). These forms were further revised and tailored to the current review. One reviewer extracted data (CN), but to ensure accuracy and consistency of the process, a sample of 20% of the articles was randomly selected for extraction by a second reviewer (AO). The data extracted by the two reviewers were then compared, and no significant discrepancies or errors were detected.

Analysis

The evidence was qualitatively reviewed with studies being grouped by interventions and comparisons and summarised according to the outcomes assessed according to Kirkpatrick levels. Evidence tables detailing study characteristics (including population, intervention, comparison, outcomes and design), results and authors' conclusions are provided. We meta-analysed immediate and long-term knowledge scores. Data were combined using weighted mean differences (WMDs), inverse variance methods and random effects models. Studies were grouped by design, and meta-analysis was performed separately for randomised controlled trials (RCTs) and non-randomised studies. For the purpose of this analysis, long-term outcomes were defined as the latest examination scores reported, provided the examination was not given immediately after the teaching session. Those that were given immediately following the teaching session were designated as immediate knowledge score outcomes. Heterogeneity was quantified using the I^2 statistic; an I^2 value of greater than 50% was considered substantial heterogeneity (Higgins & Thompson 2002; Higgins et al. 2003). Knowledge scores were assessed using different scales (e.g. 0–100, 0–7, etc.); we conducted sensitivity analyses using standardised mean differences to account for this variability. Analyses were conducted using RevMan 5.0 (The Cochrane Collaboration, Copenhagen, Denmark). Results are reported with 95% confidence intervals (CIs) and statistical significant was set at $p < 0.05$.

Results

Figure 1 presents a flow diagram of the study selection process. Eight hundred and fourteen studies were identified by electronic database searches, and 193 studies were identified by reference and hand searches. Of these 1007 studies, title and abstract screening identified 220 potentially relevant studies that warranted full-text review. Authors of included studies

Table 4. Inclusion and exclusion criteria applied to potentially relevant studies to determine suitability for systematic review purposes.

	Inclusion criteria	Exclusion criteria
Population	Medical students Residents Physicians Nursing students/nurses Pharmacy students/pharmacists Dental students/dentists Veterinary medicine Trainees/veterinarians Dietician trainees/dieticians Clinical psychology trainees/Clinical psychologists Other allied health professionals	Non-health professions trainees
Intervention	Audience response in conjunction with: Lectures Workshops Small group learning sessions Clinical teaching Videos Other teaching sessions	Shadowing Mentoring Practice audits Feedback alone
Comparator	Any teaching method described under the inclusion criteria for 'Intervention' section without audience response. Any 'standard curriculum' without audience response	
Outcome (Based on modified Kirkpatrick's 1967 model of hierarchical outcomes)	Change in patients' health Change in behaviour Inclusion of skill in clinical practice Change in skills OSCE scores Observed assessment scores Change in knowledge Written exam scores Change in attitudes/perceptions Confidence self ratings Comfort self ratings Learner reaction Satisfaction with teaching method Satisfaction with instructor	
Study type	Comparative studies, which provide primary data for any of the outcomes listed above, including the following designs: Randomised controlled trials Non-randomised control trials Cohort studies Controlled before and after studies Interrupted time series Other robust comparative studies English language (Morrison et al. 2009)	Studies reporting on needs assessments for audience response systems Studies reporting the prevalence of audience response systems Opinion papers Articles not in the English language

were contacted by email, and this yielded six additional studies giving a total of 1013 studies for review. Inclusion criteria were applied to the full text of these 226 studies. As a result, 21 studies met inclusion criteria for this review.

Among the included studies, nine were RCTs (Miller et al. 2003; Palmer et al. 2005; Pradhan et al. 2005; Duggan et al. 2007; Plant 2007; Elashvili et al. 2008; Rubio et al. 2008; Liu et al. 2010; Moser et al. 2010), two were non-randomised controlled trials (NRCTs) (Schackow et al. 2004; Patterson et al. 2010), two were prospective cohort studies (O'Brien et al. 2006; Stein et al. 2006) and eight were non-concurrent cohort studies (Halloran 1995; Slain et al. 2004; Barbour 2008; Berry 2009; Cain et al. 2009; Doucet et al. 2009; Lymn & Mostyn 2009; Grimes et al. 2010).

Most of the studies were conducted in the United States (16 studies; Halloran 1995; Miller et al. 2003; Schackow et al. 2004; Slain et al. 2004; Pradhan et al. 2005; O'Brien et al. 2006; Stein et al. 2006; Plant 2007; Elashvili et al. 2008; Rubio et al. 2008; Berry 2009; Cain et al. 2009; Grimes et al. 2010; Liu 2010; Moser et al. 2010; Patterson et al. 2010) with the remainder based in the United Kingdom (Barbour 2008; Lymn & Mostyn 2009), Australia (two studies; Palmer et al. 2005; Duggan et al. 2007) and Canada (Doucet et al. 2009). Thirteen of the 21 studies were concerned with undergraduate health professions education including four studies in nursing (Halloran 1995; Stein et al. 2006; Berry 2009; Patterson et al. 2010), three studies in medicine (Palmer et al. 2005; Duggan et al. 2007; Moser et al. 2010), two studies in dentistry (Barbour 2008; Elashvili et al. 2008), two studies in pharmacy (Cain et al. 2009; Liu et al. 2010) and two studies in veterinary medicine (Plant 2007; Doucet et al. 2009). Three studies involved medical residents (Palmer et al. 2005; Pradhan et al. 2005; Rubio et al. 2008). Three studies involved graduate trainees, two in pharmacy (Slain et al. 2004; Moser et al. 2010) and the other in nursing (Grimes et al. 2010). Practicing professionals were the subjects in two studies, one involving physicians

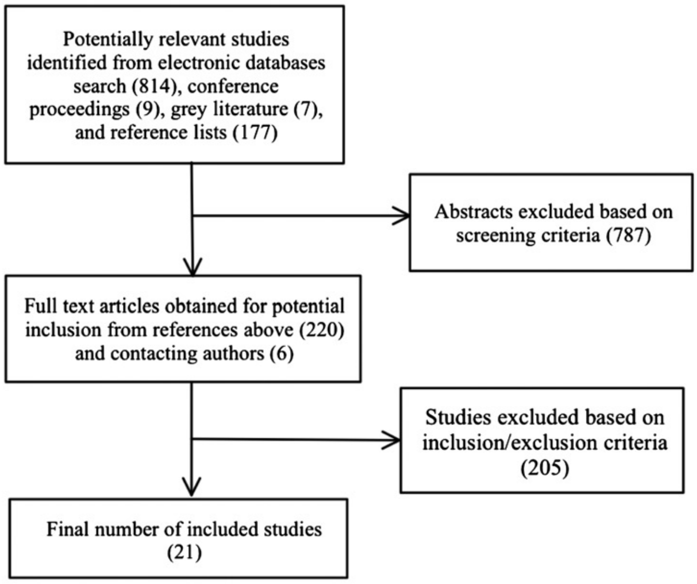

Figure 1. Flow diagram of included studies.

(Miller et al. 2003) and the other nurses (Lymn & Mostyn 2009). Several studies assessed more than one level of Kirkpatrick learning outcomes. All 21 studies assessed change in knowledge, and six studies assessed a change in learner reactions (Miller et al. 2003; Slain et al. 2004; Duggan et al. 2007; Elashvili et al. 2008; Cain et al. 2009; Doucet et al. 2009). One of the studies assessed change in self-confidence (Doucet et al. 2009). None of the studies evaluated skills or patient outcomes. In total, 2637 participants were involved in the included studies.

Methodological quality and risk of bias of included studies

The methodological quality of the studies varied, however several weaknesses were common to particular designs. The 11 RCTs and NRCTs were assessed using the Cochrane Risk of Bias tool. The randomisation process and allocation concealment were unclear in all nine randomised control trials (Miller et al. 2003; Pradhan et al. 2005; Duggan et al. 2007; Plant 2007; Palmer & Devitt 2007; Elashvili et al. 2008; Rubio et al. 2008; Liu et al. 2010; Moser et al. 2010). Two trials were not randomised (Schackow et al. 2004; Patterson et al. 2010). In about half of the trials (Pradhan et al. 2005; Elashvili et al. 2008; Rubio et al. 2008; Liu et al. 2010; Moser et al. 2010; Patterson et al. 2010), outcome data were either incomplete or inadequately addressed. One trial (Moser et al. 2010) was found to be at risk of selective outcome reporting. Eight trials (Miller et al. 2003; Schackow et al. 2004; Pradhan et al. 2005; Duggan et al. 2007; Plant 2007; Elashvili et al. 2008; Rubio et al. 2008; Moser et al. 2010) did not present any baseline characteristics of the groups being compared, and one trial reported general baseline imbalance.

For the majority of prospective and non-concurrent cohorts (Halloran 1995; Slain et al. 2004; O'Brien et al. 2006; Stein et al. 2006; Barbour 2008; Berry 2009; Cain et al. 2009; Doucet et al. 2009; Grimes et al. 2010), the exposed and non-exposed groups were drawn from the same community, and the learners were truly representative of the average participant in the community. One non-concurrent cohort was not drawn from the same community (Lymn & Mostyn 2009). However, none of the studies took into account the comparability of cohorts or controlled for potential confounders in the association between intervention and outcomes (skills, knowledge and confidence). All of the studies had a clear definition of the outcome, and reported outcomes were based on record linkage. Three studies provided no statement regarding completeness of follow-up (Stein et al. 2006; Barbour 2008; Lymn & Mostyn 2009). One study had less than 10% of its subjects lost, and this small loss is unlikely to introduce bias (Slain et al. 2004). One study did not have adequate follow-up of participants, as its loss to follow-up rate was greater than 10% of study participants and there was an incomplete description of those lost (Doucet et al. 2009). Further detailed results of the assessments of methodological quality are available from the authors on request.

Characteristics of included studies

Table 5 provides a summary of the interventions, comparators, outcomes measured and main findings of all included studies. All studies reported knowledge as an outcome, one reported learner self-confidence (Doucet et al. 2009) and six reported learner reaction (Miller et al. 2003; Slain et al. 2004; Duggan et al. 2007; Elashvili et al. 2008; Cain et al. 2009; Doucet et al. 2009). Tables 6 and 7 detail the characteristics and results of all included studies. The following provides a narrative overview of the results grouped according to educational outcome.

Knowledge. All 21 studies, involving 2637 participants, compared knowledge-based learning outcomes between ARS lectures vs. traditional lectures (20 studies) and ARS tutorial

Table 5. Summary of findings.

Outcome	Intervention	Comparator	Findings: Any significant difference			Study Design and Number of Participants Enrolled
			<5%	5–10%	>10%	
Knowledge	Lecture with ARS	Traditional non-interactive lecture			Favours ARS	1 RCT ($n=22$)
			No difference			1 RCT ($n=127$)
					Favours ARS	1 RCT ($n=77$)
			No difference			1 RCT ($n=20$)
					Favours ARS	1 RCT ($n=17$)
					Favours ARS	1 NRCT ($n=24$)
		Traditional interactive lecture	No difference			1 RCT ($n=283$)
				Favours ARS		1 RCT ($n=179$)
				Favours ARS		1 RCT ($n=86$)
			No difference			1 NRCT ($n=70$)
			No difference			1 NCC ($n=28$)
				Favours ARS		1 NCC ($n=254$*)
					Favours ARS	3 NCC ($n=131$, $n=141$ and $n=131$)
			Favours ARS			1 NCC ($n=169$)
			No difference			1 NCC ($n=142$)
			Favours ARS			1 NCC ($n=126$)
			Favours ARS			1 Prospective cohort ($n=148$)
		Lecture, unknown interaction		Favours ARS		1 NCC ($n=88$)
				Favours ARS		1 NCC ($n=66$*)
			No difference			1 Prospective cohort ($n=283$)
	Tutorial with ARS	Standard tutorial (interactive)		Favours ARS		1 RCT ($n=102$)
Self-confidence	Lecture with ARS	Traditional interactive lecture	Favours ARS			1 NCC ($n=169$)
Reaction	Lecture with ARS	Traditional non-interactive lecture	Mixed			1 RCT ($n=127$)
			Favours ARS			1 RCT ($n=77$)
		Traditional interactive lecture	Favours ARS			1 NCC ($n=254$*)
			Favours ARS			3 NCC ($n=131$, $n=141$, $n=131$)
			Favours ARS			1 NCC ($n=169$)
			Favours ARS			1 RCT ($n=283$)

Notes: ARS = audience response system; RCT = randomised controlled trial; NRCT = non-randomised controlled trial; NCC = non-concurrent cohort.
*The exact number of participants enrolled in the study was not reported.

vs. traditional tutorial (one study). Fourteen studies reported a statistically significant difference in at least one knowledge assessment score in favour of ARS. In terms of the magnitudes of difference, of the studies with statistically significant differences, five reported a difference of at least 10% in knowledge assessment scores favouring the ARS group. Of these five studies, three were RCTs ($n=22$, $n=77$ and $n=17$; Pradhan et al. 2005; Rubio et al. 2008; Elashvili et al. 2008), one was an NRCT ($n=24$; Schackow et al. 2004) and one was a non-concurrent cohort ($n=131$). The subjects of these studies were medical residents (three studies Schackow et al. 2004; Pradhan et al. 2005; Rubio et al. 2008;), undergraduate dental students (one study; Elashvili et al. 2008) and graduate pharmacy students (one study; Slain et al. 2004). Interestingly, there were only three studies (Palmer et al. 2005; Pradhan et al. 2005; Rubio et al. 2008) in the review with medical resident participants and all three showed a greater than 10% increase in knowledge assessment scores using ARS. Six studies reported a statistically significant difference in knowledge assessment scores of at least 5% in favour of the ARS group. There were three RCTs ($n=179$, $n=102$ and $n=86$; Palmer et al. 2005; Liu et al. 2010; Moser et al. 2010) and three non-concurrent cohort studies ($n=88$, $n=66$ and $n=254$; Cain et al. 2009; Lymn & Mostyn 2009; Grimes et al. 2010). The participants varied, including undergraduate pharmacy students (two studies; Cain et al. 2009; Liu et al. 2010), undergraduate medical students (one study; Palmer et al. 2005), graduate nursing students (one study; Grimes et al. 2010), graduate pharmacy students (one study; Moser et al. 2010) and health professionals (one study; Lymn & Mostyn 2009).

Three studies reported a statistically significant difference in knowledge assessment scores that was less than 5% favouring ARS. Two of these were non-concurrent cohort studies ($n=126$ and $n=169$; Berry 2009; Doucet et al. 2009) and one was a prospective cohort study ($n=148$; O'Brien et al. 2006). These studies involved participants from undergraduate nursing (one study; Berry 2009), undergraduate medicine (one study; O'Brien et al. 2006) and undergraduate veterinary medicine (one study) programs (Doucet et al. 2009).

Seven studies reported no statistically significant difference in any knowledge assessment measure. Three of these studies were RCTs ($n=283$, $n=55$ and $n=20$; Miller et al. 2003; Duggan et al. 2007; Plant 2007), one was an NRCT ($n=70$; Patterson et al. 2010), two were non-concurrent cohort studies ($n=28$ and $n=142$; Halloran 1995; Barbour 2008) and one was a prospective cohort ($n=283$; Stein et al. 2006). Of the seven studies showing no significant difference, participants from undergraduate nursing (three studies; Halloran 1995; Stein et al. 2006; Patterson et al. 2010), undergraduate dentistry

Table 6. Study characteristics.

Citation	Institution	Design	Population	Research question/purpose	Intervention	Comparator	Primary outcomes
(Duggan et al. 2007)	The University of Adelaide, Faculty of Health Sciences	RCT	Undergraduate medical	To examine the effect an ARS when used as an integral part of a lecture, in terms of cognitive outcomes, interaction and lecturer and student satisfaction come up with their own material	ARS was used and examined in two lecture series, 3-hour blocks held every week (5–6 questions per lecture)	Traditional Lecture format	Knowledge: Immediate and 8–12 week post-quiz scores Learner reaction: Reaction to ARS and non-ARS lectures
(Elashvili et al. 2008)	University of Iowa College of Dentistry	RCT	Undergraduate dental	To evaluate student performance on written and psychomotor skill tests receiving conventional lectures versus an interactive ARS lecture	ARS was used with 12 questions asked throughout the lecture (some before and most after the concepts were taught)	Traditional lecture format with no questions	Knowledge: Immediate post-test, unit exam, final exam and practical exam scores Learner reaction: Student perception of ARS vs. non-ARS lectures
(Liu et al. 2010)	Midwestern University Chicago College of Pharmacy	RCT	Undergraduate pharmacy	To evaluate the effectiveness of an ARS on short- and long-term learning	ARS used in one lecture with five questions	A traditional lecture with the same five questions using hand raising to answer	Knowledge: Immediate and 1 month post-quiz scores
(Miller et al. 2003)	Nation-wide Study in 5 Centres (USA)	RCT	Health professionals	To examine the use of ARS as an interactive teaching tool for healthcare providers	ARS questions interspersed throughout the CRT presentation.	CRT with the same questions but with voluntary response	Learner reaction: Participant perceptions Knowledge: Immediate after quiz score
(Moser et al. 2010)	Wayne State University, Eugene Applebaum College of Pharmacy and Health Sciences	RCT	Graduate pharmacy	To evaluate the effect of utilising ARS on student understanding of material and overall class interaction	ARS was used for a 2-hour lecture with questions asked throughout	Traditional lecture with same questions being asked using voluntary participation	Knowledge: 10 Q score, exam average, block average Learner reaction: student interaction
(Palmer et al. 2005)	University of Adelaide, South Australia	RCT	Undergraduate medical	To measure the effect of the ARS on student learning and reaction within a small group tutorial	ARS used in two tutorials with multiple choice questions and multi-stage questions	Regular tutorial format with no ARS (standard questioning)	Knowledge: Pre-test, immediate and 6-week post-test scores, and retention calculation on both multiple choice and multi-stage questions Learner reaction (non-comparative)
(Plant 2007)	Oregon State University College of Veterinary Medicine	RCT	Undergraduate veterinary medicine	To evaluate the potential benefits of ARS for short-term and long-term knowledge retention and on student reaction	ARS used in 5 lectures (3–5 question slides/lecture)	Traditional lecture without the question slides	Knowledge: Immediate and delayed test scores Learner reaction (non-comparative)
(Pradhan et al. 2005)	UMDNJ-Robert Wood Johnson Medical School	RCT	Medical residents	To compare delivery methods of lecture material regarding contraceptive options by either traditional or interactive ARS lecture style	ARS used in one resident lecture on contraceptive options	Traditional didactic lecture with no interaction	Knowledge: Pre-test and post-test scores Learner reaction (non-comparative)
(Rubio et al. 2008)	Department of Radiology, Cincinnati Children's Hospital	RCT	Medical residents	To develop techniques for optimising educational time by evaluating the effect of ARS on residents' retention of material in both the short and long term	ARS used in one lunchtime lecture (five questions asked)	Standard didactic format (no questions or interaction)	Knowledge: Immediate and 3-month post–lecture test scores Learner reaction (non-comparative)

(Patterson et al. 2009)	Widener University, School of Nursing	NRCT	Undergraduate nursing	To evaluate the effectiveness of a teaching intervention using an ARS in the classroom by examining learning outcomes	ARS used in a lecture with questions dispersed through lecture. It was used for six lectures throughout the semester	Traditional lecture with same questions being asked using hand raising	Knowledge: Unit and final exam scores Learner reaction (non-comparative)
(Schackow et al. 2004)	Department of Family Medicine, University of Illinois	NRCT	Medical residents	To evaluate if an ARS used during didactic lectures can improve learning outcomes by family medicine residents and to identified factors influencing ARS assisted learning outcomes	ARS used in monthly lectures (seven additional ARS multiple choice question slides in the presentation)	Traditional lecture format in an interactive (with 7 questions) and non-interactive (no questions) format	Knowledge: Immediate post-lecture quiz scores, 1-month post-lecture quiz scores
(Barbour 2008)	University of Bristol Dental School	Non-concurrent cohort	Undergraduate dental	To assess the student perception of ARS implementation and the impact on end-of-course examination results	ARS was used for 9 lectures and 1 large tutorial session (between 6 and 15 questions/lecture)	Traditional lecture and tutorial with standard questions	Knowledge: Final exam scores Student reaction (non-comparative)
(Berry 2009)	University of Wisconsin-Eau Claire College of Nursing and Health Sciences	Non-concurrent cohort	Undergraduate nursing	To assess the effect of ARS on exams scores, final grades and the level of participation	ARS used in lecture throughout the semester in class and at remote locations (10 questions/lecture)	Traditional lecture with pre-class quizzes (10 questions/quiz)	Knowledge: Unit exam and final exam scores, overall grades Learner reaction (non-comparative)
(Cain et al. 2009)	College of Pharmacy, University of Kentucky	Non-concurrent cohort	Undergraduate pharmacy	To see if ARS could maintain student attention, improve student comprehension, and improve student grades and reactions	ARS used during lecture (ARS used for first 28 of 42 lectures followed by 14 lectures of traditional classroom lecture)	Traditional lecture format with oral questioning	Knowledge: Final grades of class and personal averages compared with previous year Learner reaction: Overall course and instructor rating
(Doucet et al. 2009)	University of Montreal's College of Veterinary Medicine	Non-concurrent cohort	Undergraduate veterinary medicine	To evaluate whether the use of an ARS would promote an active learning environment during case-based discussions in large groups, have an impact on student motivation and improve long-term retention	ARS was used in a 2 hour case-based discussion lectures at the of the five-week blocks (three blocks)	Case-based discussion groups with open-ended voluntary participation questions	Knowledge: Final exam scores, and 1 year post-test scores Learner reaction: Student perceptions and preparation

(continued)

Table 6. Continued.

Citation	Institution	Design	Population	Research question/purpose	Intervention	Comparator	Primary outcomes
(Grimes et al. 2010)	Austin School of Nursing, University of Texas	Non-concurrent cohort	Graduate nursing	To evaluate the usefulness of integrating ARS technology into an accelerated graduate nursing program by measuring leaner and instructor satisfaction and examining student achievement	ARS used in a weekly 2-hour lecture course for 15 weeks (up to eight questions embedded in lectures)	Traditional lecture format	Knowledge: Standardised final exam scores Learner reaction (non-comparative)
(Halloran 1995)	Nursing Department, Western Connecticut State University	Non-concurrent cohort	Undergraduate nursing	To assess the effect of computer aided instruction and keypad questions compared with traditional classroom lecture on student achievement	Computer-aided and ARS incorporated lectures (used throughout semester in most or all lectures)	Traditional lecture format (overhead, discussion and oral questioning)	Knowledge: Two midterm exams and final exam scores Learner reaction (non-comparative)
(Lymn & Mostyn 2009)	University of Nottingham School of Nursing	Non-concurrent cohort	Health professionals	To incorporate and evaluate the use of an ARS in promoting understanding and improving student performance	ARS use in lecture (127 questions were incorporated into eight lectures with 13 questions repeated two times or more)	Traditional lecture without the ARS	Knowledge: Formative and summative test scores Learner reaction (non-comparative)
(Slain et al. 2004)	School of Pharmacy, West Virginia University	Non-concurrent cohort	Graduate pharmacy	To evaluate the impact of interactive ARS on student learning, interest and satisfaction.	ARS used during lecture in three separate courses (43, 127 and 70 ARS questions used in each course, respectively, throughout the semester)	Traditional lecture format with mini case questions	Knowledge: Exam scores, final grades and specific question subsets Learner reaction: Multi-question survey
(O'Brien et al. 2006)	Case Western Reserve University (Case) School of Medicine	Prospective cohort	Undergraduate medical	To assess the impact of ARS on final exam scores and student reaction	ARS was used throughout the semester (17 of 31 large group lectures and 3–4 questions/lecture asked at the end of lecture)	The same class and questions as intervention, but those students who did not once use respond to the ARS questions	Knowledge: Final exam scores Learner reaction (non-comparative)
(Stein et al. 2006)	University of Kentucky College of Medicine	Prospective cohort	Undergraduate nursing	To determine whether the use of ARS for pre-test reviews improved student learning outcomes, to describe the steps involved in creating an ARS review, and to encourage nurse educators to implement this technology in their classrooms.	There were six of eight pre-test reviews given using the ARS. It was used in conjunction with a Jeopardy style game.	Two of eight pre-test reviews were given in the traditional review format (lecture-style overview of notes)	Learner reaction (non-comparative) Knowledge: Test scores

Table 7. Main findings of the review.

Citation	Design	Outcomes	Qualitative results	Author's conclusions	Stated limitations
(Duggan et al. 2007)	RCT	Knowledge: Immediate and 8–12-week post-quiz scores Student reaction: Attitudes of ARS and non-ARS lectures	Knowledge sores were not completely reported, however, the author reported was no difference in MCQ scores between ARS and traditional lectures ($p = 0.785$). The cervical cancer lectures showed higher student ranking in favour of ARS in all parameters. The breast cancer lectures showed higher ranking in favour of traditional lectures in five of seven parameters ($p < 0.001$)	'In this setting, EVS [ARS] technology used in large group lectures did not offer significant advantages over the more traditional lecture format'	There was poor participation, small sample size; only two lectures on two topics were examined
(Elashvili et al. 2008)	RCT	Knowledge: Immediate post-test, unit exam, final exam and practical exam scores Student Reaction: Student perception of ARS vs. non-ARS lectures	Statistically significant differences were found in favour of the ARS in scores on the immediate written post-test (mean scores 8.7/10 vs. 7.6/10, $p = 0.002$) and in performance bond strength testing (means of 26.7/40 vs. 23.3/40, $p = 0.039$) for the lecture 'Principles of Dental Bonding'. The other examinations/skill testing showed no significant difference between the two groups. The responses to the question 'I can easily transfer my knowledge gained from the class to the practical examination' were significantly higher in the ARS group receiving the 'Principles of Dental Bonding' lecture. All other questions showed no difference	'This technology has the potential to increase student knowledge retention and the ability to transfer the student's knowledge through psychomotor skill performance when used carefully in the context of the lecture'. These results indicate that ARS is a promising teaching tool for dental education	One lecture (with ARS) was given in the afternoon on the same day of a major exam in the morning
(Liu et al. 2010)	RCT	Knowledge: Immediate and 1-month post-quiz scores	Students who attended the ARS class scored an average 1 point higher on the immediate quiz than students who were assigned to the control group (10.7/16 vs. 9.7/16, $p = 0.02$). No significant difference was seen between the quiz 2 scores of the two groups (9.5/16 vs. 9.5/16, $p = 0.99$)	'The use of a student response system can positively impact students' short-term learning; however, that positive effect did not appear to last over time'	The study was non-concurrent, allowing possible communication between two groups. The small number of questions in the quiz may not allow for significant difference to be shown
(Miller et al. 2003)	RCT	Student Reaction: Participant perceptions Knowledge: Immediate after quiz score	ARS participants rated the quality of the presentation, the quality of the speaker and their level of attention significantly higher than non-ARS participants ($p < 0.05$). Knowledge scores were not significantly different between the two groups (3.9/5 vs. 4.3/5, $p = 0.129$)	The overall opinion of the ARS was favourable; however, the comparative test assessment showed no difference between groups	The study relied on self-reported data, it was not a blinded study and there was lecturer variability
(Moser et al. 2010)	RCT	Knowledge: Subset of ARS questions, exam average, overall block grade and average GPA Student reaction (non-comparative)	Statistically significant differences were found in the 10 exam questions that covered material from the 2 hours of ARS/non-ARS lecture (6.6/10 vs. 5.8/10, $p = 0.03$) and overall block grade (82.9% vs. 80.4%, $p = 0.047$) favouring ARS lectures. No difference was found in the exam average (78% vs. 76.1%, $p = 0.31$) or the average GPA (3.0 vs. 3.0, $p = 0.409$)	Use of the ARS resulted in better exam scores and increased in-class interaction between students and instructor. These results suggest that the ARS is an effective tool at promoting learning and should be utilised throughout the pharmacy curriculum	NS
(Palmer et al. 2005)	RCT	Knowledge: Pre-test, immediate post-test, 6-week post-test scores, and retention calculation on both multiple choice and multi-stage questions Student reaction (non-comparative)	Different quiz question types were analysed for ARS and non-ARS tutorials. All pre-test scores were statistically similar. For multiple-choice questions (MCQ) in the GI haemorrhage tutorial no scores were significantly different. For MCQ in the abdominal pain tutorial immediate quiz scores (7.60/11 vs. 6.94/11, $p = 0.03$) favoured traditional tutorials, whereas all other scores were not significantly different. For Multistage questions (MSQ) in the GI haemorrhage tutorial, both the 6-week post-quiz (5.96/12 vs. 6.72/12, $p = 0.01$) and retention scores (−0.28 vs. 0.52, $p = 0.01$) favoured the ARS tutorial. For MSQ in the abdominal pain tutorial immediate quiz scores (6.79/12 vs. 6.22/12, $p = 0.03$) favoured traditional tutorials, while retention scores (−0.87 vs. −0.02, $p = 0.02$) favoured ARS tutorials	'Electronic voting systems can provide a stimulating learning environment for students and in a small group tutorial may improve educational outcomes'	There was a limited amount of ARS equipment (enough for 30 students)

(continued)

Table 7. Continued.

Citation	Design	Outcomes	Qualitative results	Author's conclusions	Stated limitations
(Plant 2007)	RCT	Knowledge: Immediate and delayed test scores Student reaction (non-comparative)	The mean short-term knowledge-retention test scores of the ARS group and non-ARS group were 81% and 78% ($p = 0.32$,), respectively. The mean long-term knowledge-retention test scores of the ARS and non-ARS groups were 54% and 55% ($p = 0.77$), respectively. The differences between groups were not significant for either time period	'Although benefits to short-term and long-term knowledge retention were not detected in this pilot study, all students responding to the survey perceived a benefit and supported the use of ARS in the clinical veterinary medicine curriculum'	The study had a very small sample size, voluntary participation and participants not blinded to the outcome (after lecture test)
(Pradhan et al. 2005)	RCT	Knowledge: Pre-test and post-test scores Student reaction (non-comparative).	Residents who received ARS interactive lectures showed a 21% improvement (78%–95%) between pre-test and post-test scores, whereas residents who received the standard lecture demonstrated a 2% improvement (80%–82%). There was a significant difference in improvement of test scores favouring the ARS lectures ($p = 0.018$).	The evidence shows the effectiveness of the audience response system for knowledge retention, which suggests that it may be an efficient teaching tool for residency education.	It is possible that the effect seen was due to the novelty of the system and an increase in attention.
(Rubio et al. 2008)	RCT	Knowledge: Immediate and 3-month post-lecture test scores Student reaction (non-comparative)	Immediate post-test scores (76.4% vs. 60.0%, $p = 0.02$) significantly favoured ARS lectures over traditional lectures. Three-month post-scores (58.2% vs. 27.5%, $p < 0.001$) also significantly favoured ARS lectures	The ARS is an effective alternative to either traditional lecture or to the hot seat style. Most resident teaching use one of these two styles	The sample size was small in this study, and there were limitations that did not allow every initial learner to be retested
(Patterson et al. 2009)	NRCT	Knowledge: Unit and final exam scores Student reaction (non-comparative)	Comparing scores after ARS vs. traditional lectures; unit 1 exam scores (43.76/50 vs. 44.16/50, $p = 0.562$), unit 2 exam scores (40.42/50 vs. 41.19/50, $p = 0.332$), unit 3 exam scores (39.18/50 vs. 37.53/50, $p = 0.060$), and final exam scores (65.84/80 vs. 64.59/80, $p = 0.340$), there was no statistical difference in any test scores	There is no observed difference in achievement on exams but this may be due to limited exposure (six lectures for 15 minutes each). The study was promising because it showed increased student engagement	The study was a convenience sample. There was a limited exposure and technology implementation problem that may have affected the outcomes
(Schackow et al. 2004)	NRCT	Knowledge: Immediate post-lecture quiz scores, 1-month post-lecture quiz scores	Immediate post-lecture quiz scores were 4.25/7 with non-interactive lectures, 6.50/7 following interactive lectures without ARS and 6.70/7 following ARS lectures. The difference in scores following ARS or interactive lectures versus non-interactive lectures was significant for both ($p < 0.001$). Six-week post-lecture quiz scores were 3.39/7 with non-interactive lectures, 4.22/7 following interactive lectures and 4.67/7 following ARS lectures. ARS was significantly higher than basic ($p < 0.05$), whereas interactive was not ($p = 0.11$). The difference between the ARS and interactive lectures was not significant for either quiz	'Both audience interaction and ARS equipment were associated with improved learning outcomes following lectures to family medicine residents'	This was a small study, it was not randomised and there was not 100% attendance of participants. The study was limited to family medicine residents only
(Barbour 2008)	Non-concurrent cohort	Knowledge: Final exam scores Student reaction (non-comparative)	There was no significant difference in final exam scores (71.60% vs. 69.90%, $p = 0.44$) comparing ARS lectures vs. traditional lectures	The ARS system proved very popular with the students. There was, however, no statistically significant impact on the results of the examination at the end of the course	This was not a controlled experimental study, and the sample size was limited

(Berry 2009)	Non-concurrent cohort	Knowledge: Unit exam and final exam scores, overall grades Student reaction (non-comparative)	Statistically significant differences were found in Exam 2 (91.23% vs. 86.93%, $p = 0.000$) and Corse Grades (95.03% vs. 93.33%, $p = 0.000$) in favour of ARS lectures. No difference was found in Exam 1 (90.57% vs. 90.48%, $p = 0.880$) or the Final exam (90.78 vs. 89.80%, $p = 0.180$)	Knowledge outcomes showed that some increase when the ARS was used. Student satisfaction supported the use of ARS as a way to engage students	Control group, although similar in admission GPA and size, were not matched for other characteristics. The study was in a rural nursing program and may not be applicable on a larger scale
(Cain et al. 2009)	Non-concurrent cohort	Knowledge: Final grades of class and personal averages compared with previous year Student reaction: Overall course and instructor rating	The final average for the ARS cohort (2008 class) was 89.9% while the comparative cohorts were 86.1% (2007, $p < 0.05$) and 81.8% (2006, $p < 0.001$). This was a significant difference favouring ARS above either cohort. Overall ARS course ratings were 3.3/4.0 vs. 2.4/4.0 for comparator. Overall ARS instructor ratings were 3.5/4.0 vs. 2.65/4.0 for comparator. The author stated significantly higher ratings favouring ARS in both categories but no p value was given	While using the ARS strategy, 'classroom attendance improved, course grades increased, and course/ instructor ratings improved'. A causal relationship cannot be established but this shows positive results for ARS use	This was not a controlled experimental study
(Doucet et al. 2009)	Non-concurrent cohort	Knowledge: Final exam scores, and 1-year post-test scores Student reaction: Student perceptions and preparation.	Final examination results (92.2% vs. 89.0%, $p = 0.03$) were significantly greater for the ARS lectures. One-year post-tests showed no significant difference. Student confidence with difficult skills was significantly higher for the ARS group in three of six categories of questions and do different in the other three. Student reaction results indicated that the use of an ARS provided an active learning environment by favouring engagement, observation and critical reflection and by increasing student and teacher motivation.	ARS use significantly improved the learning experience associated with case-based discussions in a large group of undergraduate students.	The study was not randomised; there was no blinding or crossing over. There was a different female to male ratio in to two cohorts.
(Grimes et al. 2010)	Non-concurrent cohort	Knowledge: Standardised final exam scores Student reaction (non-comparative)	Standardised final exam average (89.23% vs. 80.79%, $p < 0.001$) significantly favoured the ARS lectures over traditional lecture	Student and faculty satisfaction surveys as well as student performance outcomes all reflected positively on the use of the ARS	No comparative analysis of group demographics was done, nor was a pre-test–post-test design used

(continued)

Table 7. Continued.

Citation	Design	Outcomes	Qualitative results	Author's conclusions	Stated limitations
(Halloran 1995)	Non-concurrent cohort	Knowledge: Two midterm exams and final exam scores Student Reaction (non-comparative)	There was no significant difference in class averages for the ARS group vs. comparator for Midterm 1 (76.7% vs. 82.1%), Midterm 2 (80.1% vs. 82.6%), or the Final exam (83.4% vs. 78.4%). No p values were given	The ARS with computer-aided help is a viable option in nursing education. Test scores were statistically the same, however a trend toward improvement is seen in the ARS group	The study was small and was not randomised
(Lymn & Mostyn 2009)	Non-concurrent cohort	Knowledge: Formative and summative test scores Student reaction (non-comparative)	Students who had experienced ARS teaching scored significantly higher ($p < 0.05$) in both the formative and summative exam in comparison to students from cohort two. However, when compared with cohort one, there was no difference in summative or formative grades, suggesting a cohort effect. Exact scores were not reported numerically but were displayed graphically	The ARS has an overall positive impact on learning based on exam results and student opinion	There is a possible cohort affect between the experimental and one control group
(Slain et al. 2004)	Non-concurrent cohort	Knowledge: Exam scores, final grades, and specific question subsets Student reaction: Multiquestion survey	Three separate courses using ARS were examined. Students using the ARS had better scores on the Clinical Pharmacokinetics examination questions (mean scores, 82.6% vs. 63.8%, $p < 0.001$), on the cumulative final examination for Medical Literature Evaluation (82.9% vs. 78.0%, $p = 0.016$), and on the evaluable 'analysis type' examination questions in the Pathophysiology and Therapeutics course (82.5% vs. 77.4%, $p = 0.0002$). All other analysis of knowledge scores were not different. Students using the ISR system in all three courses were positive about the system	The ARS was a useful tool to encourage active student learning was well received by students. It is effective in gauging student under-standing and can positively affect student performance	This was not a controlled experimental study
(O'Brien et al. 2006)	Prospective cohort	Knowledge: Final exam scores Student reaction (non-comparative)	Mean course exam score was 81.9% for non-participants and 85.8% for the students who used the ARS at least once. This mean progressively increased to 94.4% for the students who used the system the most. A regression analysis showed a significant ($p < 0.01$) relationship between the level of participation with the ARS and exam performance. There was no significant relationship ($p > 0.1$) on the Year 1 Comprehensive Exam, arguing against a self-selection bias for better exam performers	'The use of an ARS with lec-tures for medical students seemed to improve exam performance and promote active learning'	There were a small number of participants, and only 14 students took 10 or more quizzes
(Stein et al. 2006)	Prospective cohort	Student reaction (non-com-parative) Knowledge: Test scores	Average scores of the class on examinations preceded by the interactive ARS review were compared with those on examinations preceded by the more traditional, lecture-style review; no significant improvement because of ARS use was found. There were no scores reported in the study	Class examination scores did not show a significant increase when compared with standard reviews but student opinion of the ARS was very favourable	NS

Note: NS – None stated.

(one study; Barbour 2008), undergraduate veterinary medicine (one study; Plant 2007), undergraduate medicine (one study; Duggan et al. 2007) and practicing professionals (one study; Miller et al. 2003) were involved.

The effect of ARS on short- and long-term knowledge assessment scores was examined. Nine studies examined scores from tests, quizzes or questionnaires that immediately followed exposure to ARS (Miller et al. 2003; Schackow et al. 2004; Palmer et al. 2005; Duggan et al. 2007; Plant 2007; Elashvili et al. 2008; Rubio et al. 2008; Liu et al. 2010; Moser et al. 2010). The range of number of immediate knowledge assessments performed in each of these studies was one to two. Four studies (Schackow et al. 2004; Elashvili et al. 2008; Rubio et al. 2008; Moser et al. 2010) reported a significant difference in at least one knowledge assessment score favouring ARS lectures, four (Miller et al. 2003; Palmer et al. 2005; Duggan et al. 2007; Plant 2007) reported no difference and one (Liu et al. 2010) reported immediate quiz scores favouring traditional lectures, but this difference did not extend to the long-term scores in this study.

Eighteen studies reported long-term knowledge assessment scores (at least one month later) from quizzes, tests, unit exams, final exams, class averages or overall grade point averages. The range of number of long-term knowledge assessments performed in each of these studies was one to three. Of these 18 studies, eight (Schackow et al. 2004; Slain et al. 2004; Palmer et al. 2005; Pradhan et al. 2005; Rubio et al. 2008; Cain et al. 2009; Grimes et al. 2010; Moser et al. 2010) reported a significant difference in at least one knowledge assessment score favouring ARS. The other 10 studies (Halloran 1995; O'Brien et al. 2006; Duggan et al. 2007; Plant 2007; Barbour 2008; Elashvili et al. 2008; Berry 2009; Doucet et al. 2009; Liu et al. 2010; Patterson et al. 2010) reported no difference in any score. There were no long-term knowledge assessment scores that significantly favoured traditional teaching.

Comparison group. A difference in knowledge assessment scores can have as much to do with the comparator group as with the intervention group. In order to better understand the impact of ARS on knowledge-based scores, the comparator groups were also analysed. As part of the data extraction, comparator groups were divided into interactive vs. non-interactive categories. An interactive comparator was defined as one where any similar questions were asked or any attempted interaction was observed. Six of the 21 studies compared ARS lectures with traditional lectures that were not interactive (Schackow et al. 2004; Pradhan et al. 2005; Duggan et al. 2007; Plant 2007; Elashvili et al. 2008; Rubio et al. 2008). Of these six studies, four reported a statistically significant difference in knowledge assessment scores favouring ARS and the difference in all four studies was 10% or greater (Schackow et al. 2004; Pradhan et al. 2005; Elashvili et al. 2008; Rubio et al. 2008). Eleven of the 21 studies compared ARS lectures (10 studies; Halloran 1995; Miller et al. 2003; Slain et al. 2004; Barbour 2008; Berry 2009; Cain et al. 2009; Doucet et al. 2009; Liu et al. 2010; Moser et al. 2010; Patterson et al. 2010) and tutorials (one study; Doucet et al. 2009) with traditional lectures/tutorials that were interactive. Seven of the 11 studies (Slain et al. 2004; O'Brien et al. 2006; Elashvili et al. 2008; Berry 2009; Cain et al. 2009; Doucet et al. 2009; Liu et al. 2010) reported a statistically significant difference in knowledge assessment scores. Of these seven studies, only one (Slain et al. 2004) reported a statistically significant increase of 10% or greater. Three studies did not make clear the level of interaction of the comparator. Two of these studies (Lymn & Mostyn 2009; Grimes et al. 2010) favoured ARS, while one (Stein et al. 2006) reported no difference in knowledge assessment scores. Thus, while ARS can increase knowledge-based scores, the greatest effect is seen when they are compared to non-interactive lectures.

Meta-analysis. Meta-analyses were performed for immediate and long-term knowledge outcomes. The results are shown in Figures 2 and 3, respectively. The RCTs showed no significant difference between groups in either immediate (WMD; 4.53, 95% CI −0.68, 9.74, $n=8$) or long-term (WMD 1.36, 95% CI −3.77, 6.50, $n=6$) knowledge scores. The non-randomised studies demonstrated a significant difference favouring ARS for both immediate (WMD 4.57, 95% CI 1.47, 7.67, $n=10$) and long-term (WMD 35, 95% CI 26.4, 43.6, $n=1$) knowledge scores; however, the latter analysis was based on only one study. Statistical heterogeneity was high in all groups with I^2 values ranging from 70% to 89%. There was substantial variation between studies that may contribute to the statistical heterogeneity observed; this includes differences in characteristics of the participants (e.g. professional groups, undergraduate vs. other), content of the lectures, comparison groups (i.e. interactive vs. non-interactive comparators), individuals delivering the lectures, methods and time points for outcome assessment, as well as other study design features (e.g. concurrent vs. non-concurrent controls).

We conducted sensitivity analyses using standardised mean differences to account for the variation in total scores used across studies. The patterns were similar to results based on WMDs with the RCTs showing no significant differences and the non-randomised studies showing significant differences of similar magnitude for both immediate and long-term knowledge scores (data not shown; available from authors on request).

Student self-confidence and learner reaction. One non-concurrent cohort ($n=169$; Doucet et al. 2009) involving undergraduate veterinary medicine students compared students' self-confidence in skills relating to clinical pharmacology after ARS and traditional instruction. The study favoured ARS lectures with self-confidence in three of six skills categories rated significantly higher by ARS participants. The other three skill categories showed no significant difference in self-confidence between ARS and traditional lecture cohorts.

Six studies involving 1236 participants compared learner reactions to the ARS enhanced teaching sessions and traditional teaching sessions. Three of the six studies were non-concurrent cohort studies (Slain et al. 2004; Cain et al. 2009; Doucet et al. 2009), whereas the other three were RCTs (Miller et al. 2003; Duggan et al. 2007; Elashvili et al. 2008). One of these studies (non-concurrent cohort; Slain et al. 2004), examined student reaction in three separate courses

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
1.2.1 RCT								
Duggan	62.3	13	29	61.6	14	26	15.1%	0.70 [−6.47, 7.87]
Elashvili	80.517	12.657	77	77.834	13.25	73	18.6%	2.68 [−1.47, 6.83]
Liu	59.375	15.625	88	59.375	18.75	91	17.6%	0.00 [−5.05, 5.05]
Moser	66	45	43	58	16	43	8.2%	8.00 [−6.28, 22.28]
Palmer	54.95	13.41	43	57.86	15.5	43	16.4%	−2.91 [−9.04, 3.22]
Plant	54	20	10	55	20	10	6.2%	-1.00 [−18.53, 16.53]
Pradhan	95	10.667	8	82	15.46	9	9.5%	13.00 [0.48, 25.52]
Rubio	58.2	19.4	11	27.5	11.6	8	8.4%	30.70 [16.70, 44.70]
Subtotal (95% CI)			**309**			**303**	**100.0%**	**4.53 [−0.68, 9.74]**
Heterogeneity: $Tau^2 = 33.48$; $Chi^2 = 23.16$, $df = 7$ ($P = 0.002$); $I^2 = 70\%$								
Test for overall effect: $Z = 1.70$ ($P = 0.09$)								
1.2.2 Other Design								
Barbour	71.6	13	75	69.9	13	67	9.9%	1.70 [−2.58, 5.98]
Berry	90.78	4.28	65	89.8	3.85	61	11.8%	0.98 [−0.44, 2.40]
Cain	89.9	15.5	127	86.1	15.5	127	10.3%	3.80 [−0.01, 7.61]
Doucet	51.3	10	86	51	8.5	83	11.1%	0.30 [−2.49, 3.09]
Grimes	89.25	4.94	49	80.79	6.6	39	11.3%	8.46 [5.97, 10.95]
Halloran	81.05	13.86	14	80.02	13.86	14	5.2%	1.03 [−9.24, 11.30]
O'Brien	85.8	13.9	78	81.9	13.9	70	9.7%	3.90 [−0.59, 8.39]
Patterson	82.3	6.19	38	80.7	7.4	32	10.7%	1.60 [−1.63, 4.83]
Schackow	66.714	6.43	12	48.429	4.71	18	9.9%	18.28 [14.05, 22.52]
Slain	82.9	11.5	75	78	12.2	66	10.2%	4.90 [0.97, 8.83]
Subtotal (95% CI)			**619**			**577**	**100.0%**	**4.57 [1.47, 7.67]**
Heterogeneity: $Tau^2 = 20.63$; $Chi^2 = 81.73$, $df = 9$ ($P < 0.00001$); $I^2 = 89\%$								
Test for overall effect: $Z = 2.89$ ($P = 0.004$)								

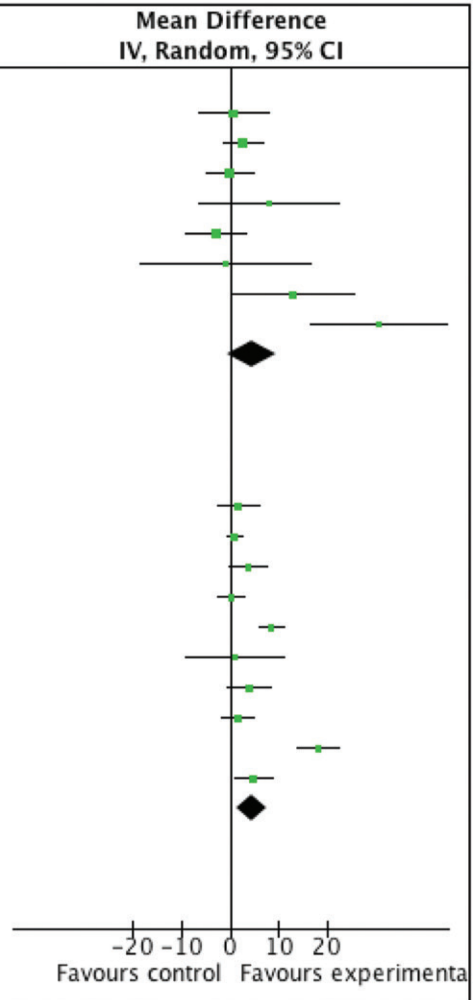

Figure 2. Meta-analysis of immediate knowledge-based test scores comparing ARS lectures (experimental) to traditional lectures (control).

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
1.1.1 RCT								
Elashvili	84.415	12.717	77	82.192	16.017	73	21.1%	2.22 [−2.42, 6.87]
Liu	66.875	19.375	88	60.625	17.5	91	19.8%	6.25 [0.84, 11.66]
Miller	55.714	18.57	164	61.429	18.57	119	21.5%	-5.72 [-10.10, -1.33]
Palmer	56.95	13.11	43	60.64	12.82	43	19.7%	-3.69 [-9.17, 1.79]
Plant	81	15	10	78	14	10	9.9%	3.00 [-9.72, 15.72]
Rubio	76.4	16.9	11	60	19	11	8.0%	16.40 [1.37, 31.43]
Subtotal (95% CI)			**393**			**347**	**100.0%**	**1.36 [-3.77, 6.50]**
Heterogeneity: $Tau^2 = 26.97$; $Chi^2 = 19.06$, df = 5 (P = 0.002); $I^2 = 74\%$								
Test for overall effect: Z = 0.52 (P = 0.60)								
1.1.2 Other Design								
Schackow	95.714	8.907	23	60.714	22.627	32	100.0%	35.00 [26.36, 43.64]
Subtotal (95% CI)			**23**			**32**	**100.0%**	**35.00 [26.36, 43.64]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 7.94 (P < 0.00001)								

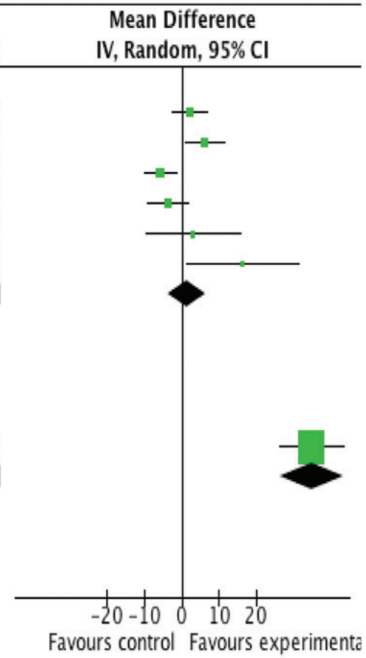

Figure 3. Meta-analysis of long-term knowledge-based test scores comparing ARS lectures (experimental) to traditional lectures (control).

($n = 131$, $n = 141$ and $n = 131$). All three of these comparator courses favoured the ARS group. In one RCT ($n = 127$; Duggan et al. 2007), the same class completed evaluations at different times. This study had mixed results in that it favoured an ARS lecture with one teacher and favoured a traditional lecture with another teacher. Two other non-concurrent cohort studies ($n = 254$ and $n = 169$; Cain et al. 2009; Doucet et al. 2009) and two RCTs ($n = 283$ and $n = 77$; Miller et al. 2003; Elashvili et al. 2008) reported student reaction that favoured the ARS. Overall, five of the six studies reported favourable learner reaction to ARS, and one study reported mixed results.

Discussion

This systematic review examined the effect of ARS on learning outcomes in health professions education. The results show some modest beneficial to neutral effects of ARS in terms of increased knowledge and self-confidence, as well as positive learner reactions. These results are reassuring for health professions educators concerned that ARS will negatively impact student achievement.

Twenty-one studies were included in the analysis and 14 of these reported statistically significant differences in favour of ARS groups over comparators in terms of knowledge scores. Five studies (Schackow et al. 2004; Slain et al. 2004; Pradhan et al. 2005; Elashvili et al. 2008; Rubio et al. 2008) demonstrated an increase of at least a 10% in knowledge assessment scores for the ARS group, an additional six studies (Palmer et al. 2005; Cain et al. 2009; Lymn & Mostyn 2009; Grimes et al. 2010; Liu et al. 2010; Moser et al. 2010) reported an increase of at least 5%, and three studies (Barbour 2008; Berry 2009; Doucet et al. 2009) reported increases of less than 5%. Only one study (Palmer et al. 2005) favoured a traditional lecture format over ARS with a statistically significant difference in scores on an immediate post-lecture quiz. However, this study reported results that favoured ARS lectures in the delayed quiz and in their analysis of knowledge retention. Thus, the effect of ARS on combined test scores was reported as favouring ARS. The authors in this study hypothesised that the findings in favour of the traditional lecture for the early quiz were due to the students' initial unfamiliarity with ARS technology. Although a number of studies reported no statistically significant difference in scores, there were no studies that reported a negative impact on knowledge-based outcome scores.

The results of our meta-analysis provide additional insights into the impact of ARS on knowledge outcomes. While the results were heterogeneous, the pooled results provide an estimate of the potential impact that ARS can have on knowledge scores. The difference for immediate knowledge showed a difference of approximately 4.5% on test scores. The magnitude of effect may be more or less depending on a number of factors, in particular, the intervention against which the ARS is compared. Through our qualitative analysis, we found that studies where ARS was compared against interactive teaching modalities showed less impact on knowledge outcomes than those that had a non-interactive comparison. Our meta-analysis also demonstrated that the magnitude of effect and statistical significance are tempered by study design: the pooled results were not significant for RCTs but were significant for the non-randomised studies. This was particularly apparent for the longer term outcomes where there was no difference among the RCTs but a substantial difference for non-randomised studies, although only one study was included; hence, we cannot make firm conclusions regarding the impact of ARS on longer term knowledge retention.

Our findings suggest that the non-randomised studies may overestimate the benefits of ARS due to methodological limitations inherent in these designs. In particular, our quality assessment highlights that many of non-randomised studies did not control for potential confounders or baseline imbalances between study groups. Future research should use randomised methods; by controlling for both known and unknown confounders between study groups, randomised studies yield less biased estimates of effect.

One non-concurrent cohort (Doucet et al. 2009) reported the self-confidence of undergraduate veterinary medicine students in clinical pharmacology. The study favoured ARS lectures; however, this single study makes it difficult to generalise these findings to other areas of education.

In terms of learner reaction, five of six studies favoured ARS lectures. As this systematic review included only comparative data, many studies that reported non-comparative student reaction were excluded. The following were three common themes noted in the review of the learner reaction data: ARS lectures were of a higher quality, they led to increased interaction and they were more enjoyable. These findings are consistent with studies that have been published describing the use of ARS in other teaching contexts (Roschelle et al. 2004; Fies & Marshall 2006; Caldwell 2007). It should be noted that for nearly all studies, ARS were novel learning tools for the students. As other authors have suggested (Caldwell 2007) some of the positive effects seen may be due to the novelty of the ARS where 'special treatment causes the improvement rather than the use of clickers'. However, this effect is difficult to assess as longer term studies have not been reported.

The current review highlights one of the caveats in interpreting this body of evidence, that is, the fact that different comparison groups were used across relevant studies. To explore the possibility of different results depending on the comparison group used, we conducted sub-group analyses to examine results of studies with interactive versus non-interactive comparators. The greatest effects on knowledge scores were seen when ARS was compared to non-interactive lectures; the differences between groups were less pronounced when non-interactive comparators were excluded. These results suggest that the positive effects of ARS on knowledge outcomes may also be produced by other interactive lecture styles or interactive modalities. These findings support previous studies that have hypothesised that increased interaction, rather than the actual technology, may be the mechanism by which ARS positively affects student achievement (Poulis et al. 1998; Caldwell 2007).

Overall, the previous reviews of ARS do not include or examine the use and impact of ARS in health professions education thoroughly nor do they systematically report the impact of the ARS on learning outcomes. The use of ARS among clinical trainees and health professionals presents a distinct work-based clinical context and has not been previously reported with similar rigour or in similar detail. For example, this is the first review to include studies of ARS in continuing professional learning. It is also the first review to explore the impact of interactive versus non-interactive comparators. Furthermore, it is the first to pool data in order to quantify the potential magnitude of effect of ARS.

In terms of limitations, inclusion bias was minimised by prospectively establishing the search strategy and by having two authors screen all potential studies, maximising the likelihood that this review is inclusive of all relevant studies. However, this review is limited by the methodological quality of included studies. Most of the studies were at a high risk of bias due to inadequate blinding of participants and/or outcome assessors. In addition, many included trials presented outcome data that was not complete or not clearly described. Either of these flaws may result in an error when estimating the intervention's effects. Similarly, few cohorts accounted for differences in learning style or level of education. Randomised trials provide a less biased comparison as the randomisation process theoretically distributes both known and unknown confounders equally between groups. We found that the magnitude of effect was smaller for randomised trials compared to non-randomised studies. Future research should aim to employ randomised methods or account for potential confounders in order to avoid overestimates of intervention effects.

Another limitation of this body of evidence is that only one study (Duggan et al. 2007) provided power calculations. Without these calculations, it is not possible to determine if observations of no difference between the interventions being compared represents actual equivalence or simply points to insufficient statistical power (i.e. type II errors). We recommend that researchers conduct sample size calculations in future studies in order to allow for more meaningful conclusions to be drawn.

The review is also limited by weaknesses inherent to the field of investigation, many of which have been previously discussed. For example, Schmidt et al. (1987) outlined the difficulty controlling for extraneous variables that may affect outcomes, particularly in studies that extend over a period of time. The authors have also detailed the struggle involved in identifying and isolating the relative contributions of different curricular components that may affect outcomes (Schmidt et al. 1987; Schmidt et al. 1996; Tamblyn et al. 2005). In addition, existing outcomes and measurement tools may ineffectively assess important areas of health professionals' competence (Berkson 1993; Vernon & Blake 1993; Distlehorst et al. 2005). This is particularly relevant to the current review as the majority of data reported focused on the lower Kirkpatrick level outcomes of knowledge scores and learner reaction.

Finally, with the heterogeneity of populations, designs, interventions, comparators and outcomes measured the findings cannot be easily generalised to health professions trainees of all levels or differing education settings. However, this review is the most comprehensive evaluation of studies pertaining to health professions in the literature and allows findings on ARS to be extended to the postgraduate and continuing professional education realms.

Conclusions

This review provides a comprehensive synthesis of the evidence to guide health professions educators regarding the implementation and use of ARS in this distinctive setting. Although causal relationships cannot be determined from this review, there were a number of interesting and novel findings. ARS did not have a consistent negative impact on student achievement in any setting or compared to any other group. However, only a few studies demonstrated large increases in knowledge scores, and these were primarily non-randomised studies that compared ARS to non-interactive teaching strategies. On further examination of the studies, comparisons of interactive teaching session to ARS lectures/tutorials revealed smaller differences favouring ARS lectures. A number of studies reported no difference in student achievement. Short-term and long-term knowledge assessment scores were affected similarly. This review also revealed an interesting trend in that all three studies examining medical residents reported a large increase in knowledge assessment scores compared to non-interactive lectures. One may hypothesise that in settings, such as medical residencies, where sleep deprivation and subsequent difficulties with attention are common and well documented, the ability of ARS to enhance learner interactivity may be even more beneficial, although further study is required.

Many health professions educators feel that the expenditure of money and time are worthwhile only if a new teaching intervention substantially impacts measurable learning outcomes. The results of this review indicate that ARS may produce improved short-term and long-term knowledge outcomes. Although ARS is not the only solution for lecturers who struggle with student engagement and poor learning outcomes, it does provide a convenient way for educators to create an interactive teaching environment. However, education programmes that already consistently use an interactive style of lecturing may not see a significant increase in knowledge scores with the implementation of an ARS. The most telling result in this review is the finding that non-randomised study designs produced more strongly positive results in favour of ARS than the higher quality randomised studies, whereas smaller if any differences in learning outcomes were seen with ARS. This in itself is a very important result that reinforces the need for curriculum planners to demand more rigorous studies prior to implementing new teaching strategies and reinforces the importance of systematic evaluations of the literature on common curricular interventions in medical education.

Acknowledgements

The authors thank Ben Vandermeer for his assistance in the statistical analysis.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Notes on contributors

CODY NELSON, BSc, is a second-year medical student and a summer studentship recipient through a Faculty Evaluation and Assessment Committee grant at the University of Alberta. He has earned a Degree with honours in Physiology and Developmental Biology.

LISA HARTLING, MSc, PhD, is an Assistant Professor, Department of Paediatrics, University of Alberta, Director of the Alberta Research Centre for Health Evidence and Director of the University of Alberta Evidence-Based Practice Center. She is a *Cochrane Collaboration* reviewer and has co-authored over 20 systematic reviews including one on problem-based learning and a recent BEME on musculoskeletal Clinical Skills.

SANDRA CAMPBELL, BA, MLS, AALIA (CP), is a Public Services Librarian and library liaison to the Faculty of Medicine and Dentistry. She is also an expert database searcher and information literacy instructor. She has co-authored and provided librarian support and consultation on a recent BEME on musculoskeletal Clinical Skills.

ANNA E. OSWALD, BMSc, MD, MMEd, FRCPC, is an Assistant Professor, consultant rheumatologist and Clinician Education for the RCPSC. She has a Master's degree in Medical Education from the University of Dundee. She has co-authored a systematic review on problem-based learning and was the team lead for a BEME on musculoskeletal Clinical Skills.

References

Abrahamson L. 2006. A brief history of networked classrooms: Effects, cases, pedagogy, and implications. In: Banks DA, editor. Audience response systems in higher education: Applications and cases. Hershey, PA: Information Science Publishing. p 25.

Alexander CJ, Crescini WM, Juskewitch JE, Lachman N, Pawlina W. 2009. Assessing the integration of audience response system technology in teaching of anatomical sciences. Anat Sci Educ 2(4):160–166.

Banks DA, Bateman S. 2004. Audience response systemsin education: Supporting a 'lost in the desert' learning scenario. International Conference on Computers in Education (ICCE2004, Acquiring and Constructing Knowledge Through Human-Computer Interaction: Creating New Visions for the Future of Learning), Melbourne.

Barbour ME. 2008. Electronic voting in dental materials education: The impact on students' attitudes and exam performance. J Dent Educ 72(9):1042–1047.

Berkson L. 1993. Problem-based learning: Have the expectations been met? Acad Med 68(10 Suppl.):79–88.

Berry J. 2009. Technology support in nursing education: Clickers in the classroom. Nurs Educ Perspect 30(5):295–298.

Cain J, Black EP, Rohr J. 2009. An audience response system strategy to improve student motivation, attention, and feedback. Am J Pharm Educ 73(2):21.

Cain J, Robinson E. 2008. A primer on audience response systems: Current applications and future considerations. Am J Pharm Educ 72(4):77.

Caldwell JE. 2007. Clickers in the large classroom: Current research and best-practice tips. CBE Life Sci Educ [Electronic Resource] 6(1):9–20.

Distlehorst LH, Dawson E, Robbs RS, Barrows HS. 2005. Problem-based learning outcomes: The glass half-full. Acad Med 80(3):294–299.

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. Med Teach 31(12):e570–e579.

Duggan PM, Palmer E, Devitt P. 2007. Electronic voting to encourage interactive lectures: A randomised trial. BMC Med Educ 7:25.

Elashvili A, Denehy GE, Dawson DV, Cunningham MA. 2008. Evaluation of an audience response system in a preclinical operative dentistry course. J Dent Educ 72(11):1296–1303.

Fies C, Marshall J. 2006. Classroom response systems: A review of the literature. J Sci Educ Technol 15(1):101–109.

Forsetlund L, Bjorndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, Davis D, Odgaard-Jensen J, Oxman AD. 2009. Continuing education meetings and workshops: Effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2:CD003030.

Grimes C, Rogers GJ, Volker D, Ramberg E. 2010. Classroom performance system use in an accelerated graduate nursing program. Comput Inform Nurs 28(2):79–85.

Halloran L. 1995. A comparison of two methods of teaching. computer managed instruction and keypad questions versus traditional classroom lecture. Comput Nurs 13(6):285–288.

Hartling L, Spooner C, Tjosvold L, Oswald A. 2010. Problem-based learning in pre-clinical medical education: 22 years of outcome research. Med Teach 32(1):28–35.

Higgins JPT, Green S. 2006. Cochrane handbook for systematic reviews of interventions. Cochrane Collaboration: Wiley & Sons, Chichester, UK.

Higgins JP, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. Stat Med. 21(11):1539–1558.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. BMJ 327(7414):557–560.

Judson E, Sawada D. 2002. Learning from past and present: Electronic response systems in college lecture halls. J Comput Math Sci Teach 21(2):167–181.

Kay RH, LeSage A. 2009. A strategic assessment of audience response systems used in higher education. Australas J Educ Technol 25(2):235–249.

Kirkpatrick DL, Kirkpatrick JD. 2006. Evaluating training programs: The four levels. 3rd ed. San Francisco, CA: Berrett-Koehler.

Liu FC, Gettig JP, Fjortoft N. 2010. Impact of a student response system on short- and long-term learning in a drug literature evaluation course. Am J Pharm Educ 74(1):6.

Lymn JS, Mostyn A. 2009. Use of audience response technology to engage non-medical prescribing students in pharmacology. ICERI2009 Proceedings, 16 to 18 November 2009 Madrid, Madrid, Spain, pp. 4518–4524.

MacArthur JR, Jones LL. 2008. A review of literature reports of clickers applicable to college chemistry classrooms. Chem Educ Res Pract 9(3):187–195.

Miller RG, Ashar BH, Getz KJ. 2003. Evaluation of an audience response system for the continuing education of health professionals. J Contin Educ Health Prof 23(2):109–115.

Morrison A, Moulton K, Clark M, Polisena J, Fiander M, Mierzwinski-Urban M, Mensinkai S, Clifford T, Hutton B. 2009. English-language restriction when conducting systematic review-based meta-analyses: Systematic review of published studies. Ottawa: Canadian Agency for Drugs and Technologies in Health.

Moser LR, Kalus JS, Brubaker C. 2010. Evaluation of an audience response system on pharmacy student test performance. Unpublished manuscript.

O'Brien TE, Wang W, Medvedev I, Wile MZ, Nosek TM. 2006. Use of a computerized audience response system in medical student teaching: Its effect on exam performance. Med Teach 28(8):736–738.

Palmer, E J, and Devitt, P G. 2007. Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper, BMC Medical Education, 2007, 7:49 [Accessed 26 April 2010] http://www.biomedcentral.com/1472-6920/7/49.

Palmer EJ, Devitt PG, De Young NJ, Morris D. 2005. Assessment of an electronic voting system within the tutorial setting: A randomised controlled trial. BMC Med Educ 5(1):24.

Patterson B, Kilpatrick J, Woebkenberg E. 2009. Evidence for teaching practice: The impact of clickers. Nurse Educ Today, 30(7):603–607.

Plant JD. 2007. Incorporating an audience response system into veterinary dermatology lectures: Effect on student knowledge retention and satisfaction. J Vet Med Educ 34(5):674–677.

Poulis J, Massen C, Robens E, Gilbert M. 1998. Physics lecturing with audience paced feedback. Am J Phys 66(5):439–441.

Pradhan A, Sparano D, Ananth CV. 2005. The influence of an audience response system on knowledge retention: An application to resident education. Am J Obstet Gynecol 193(5):1827–1830.

Roschelle J, Penuel WR, Abrahamson L. 2004. The networked classroom. Educ Leadersh 61(5):50–54.

Rubio EI, Bassignani MJ, White MA, Brant WE. 2008. Effect of an audience response system on resident learning and retention of lecture material. AJR. Am J Roentgenol 190(6):W319–W322.

Schackow TE, Chavez M, Loya L, Friedman M. 2004. Audience response system: Effect on learning in family medicine residents. Fam Med 36(7):496–504.

Schmidt HG, Dauphinee WD, Patel VL. 1987. Comparing the effects of problem-based and conventional curricula in an international sample. J Med Educ 62(4):305–315.

Schmidt HG, Machiels-Bongaerts M, Hermans H, Ten Cate TJ, Venekamp R, Boshuizen HPA. 1996. The development of diagnostic competence: Comparison of a problem-based, an integrated, and a conventional medical curriculum. Acad Med 71(6):658–664.

Simpson V, Oliver M. 2007. Electronic voting systems for lectures then and now: A comparison of research and practice. Australas J Educ Technol 23(2):187–208.

Slain D, Abate M, Hodges BM, Stamatakis MK, Wolak S. 2004. An interactive response system to promote active learning in the doctor of pharmacy curriculum. Am J Pharm Educ 68(5):1–9.

Stein PS, Challman SD, Brueckner JK. 2006. Using audience response technology for pretest reviews in an undergraduate nursing course. J Nurs Educ 45(11):469–473.

Tamblyn R, Abrahamowicz M, Dauphinee D, Girard N, Bartlett G, Grand'Maison P. 2005. Effect of a community oriented problem based learning curriculum on quality of primary care delivered by graduates: Historical cohort comparison study. BMJ 331(7523):1002.

Vernon DT, Blake RL. 1993. Does problem-based learning work? A meta-analysis of evaluative research. Acad Med 68(7):550–563.

Wells, GA, Shea, B, O'Connell, D, Peterson, J, Welch, V, Losos, M, Tugwell, P. 2011. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa: Ottawa Hospital Research Institute, [Accessed 26 April 2010] http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp