

Manuscript Title: Examining the impact of audience response systems on student performance in anatomy education: a randomized controlled trial.

Authors: Stuart J. Fergusson¹, Justine Aka², Catherine M. Hennessy³, Andrew J. Wilson⁴, Simon H. Parson⁵, Ewen M. Harrison⁶, Gabrielle M. Finn*⁷, Thomas H. Gillingwater*⁸

¹Royal Alexandra Hospital, Paisley, UK; ²Department of Anatomy, School of Medicine, Trinity College Dublin, Dublin 2, Ireland; ³Department of Anatomy, Brighton and Sussex Medical School, University of Sussex, Brighton, UK; ⁴Northern Health and Social Care Trust, Causeway Hospital, 4 Newbridge Road, Coleraine, County Londonderry.; ⁵Institute of Education for Medical and Dental Sciences, College of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK; ⁶Department of Clinical Surgery, University of Edinburgh, Edinburgh, UK; ⁷Centre for Education Development, Hull York Medical School, University of York, York, UK; ⁸Edinburgh Medical School: Biomedical Sciences (Anatomy) & Centre for Integrative Physiology, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK

*Joint senior authors

Corresponding authors: Professor T Gillingwater, University of Edinburgh, Old Medical School (Anatomy), Teviot Place, Edinburgh, EH8 9AG. Email: t.gillingwater@ed.ac.uk / Dr G Finn, University of York, Centre for Education Development, York, YO10 5DD. Email: Gabrielle.Finn@hyms.ac.uk

Word count: 2,471.

Abstract

Background and Aims: Electronic audience response systems (ARSs) offer the potential to enhance learning and improve performance. However, objective research investigating the use of ARSs in undergraduate education has so far produced mixed, inconclusive results. We investigated the impact of ARSs on short- and long-term test performance, as well as student perceptions of the educational experience, when integrated into undergraduate anatomy teaching.

Methods and Results: A cohort of 70 undergraduate medical students were randomly allocated to one of two groups. Both groups received the same anatomy lecture, but one group experienced the addition of ARSs. Multiple-choice tests were conducted before, immediately after the lecture and again 10 weeks later. Self-perceived post-lecture subject knowledge, confidence and enjoyment ratings did not differ between groups. Test performance immediately following the lecture improved when compared against baseline and was modestly but significantly superior in the group taught with ARSs (mean test score of 17.3/20 versus 15.6/20 in the control group, $p = 0.01$). Tests conducted 10 weeks after the lecture showed no difference between groups ($p = 0.61$) although overall a small improvement from the baseline test was maintained ($p = 0.02$).

Conclusions: Whilst ARSs offer opportunities to deliver novel education experiences to students, an initial superiority over standard methods does not necessarily translate into longer-term gains in student performance when employed in the context of anatomy education.

Key words: medical education; education methodology; education technology; audience response systems; anatomy.

Introduction

For many years educators have sought to improve the learning experience offered to their students by adopting new technologies. One such technology is an audience response system (ARS; sometimes referred to as electronic voting systems or “clickers”). Whilst many new technologies simply aim to update delivery methods for existing materials, such as online systems rather than books, it has been proposed that ARS approaches offer a genuinely novel way of bridging the gap between the learner and educator, particularly with respect to large group teaching [1].

Early research suggested that ARSs may increase student and teacher motivation to participate in a more active learning environment, thereby facilitating increased engagement, observation and critical reflection amongst students [2]. Current ARSs provide instantaneous and anonymous feedback of student responses, allowing students to compare their understanding with that of their colleagues [3,4]. Furthermore, as an accurate method of collecting responses, ARSs have the capacity to enhance a teacher’s ability to evaluate audience understanding in real time [5]. Potential advantages of ARSs have been reported across diverse fields of academia, including microbiology [6], business education [7], library tutorials [8], and student questionnaire feedback sessions [9].

Several lines of evidence suggest that ARS activities are popular with learners [10], and their use has been linked to improved student attendance and retention, with preliminary suggestions that they may also improve student performance [11].

ARS approaches have been used as a key component of distinct pedagogical methods such as peer instruction [12], or as an add-on to more traditional lectures [13]. However, objective data on the ability of ARSs to enhance long-term knowledge and understanding remain lacking [14]. A meta-analysis of immediate and long-term knowledge outcomes suggested that there was no difference between groups at either time-point in randomized studies, while non-randomized studies showed significant benefits for ARSs in both immediate and long-term knowledge scores [15]. As with all technologies, ARSs are not without drawbacks; they can be costly to purchase initially, require some staff training, and ARS use within teaching sessions may result in less content coverage [16].

The need for more research on the impact of ARS use in anatomy education has been clearly identified [17, 18]. Here, we have used a randomized controlled trial design to investigate whether the introduction of an ARS into medical student anatomy teaching resulted in benefits with respect to objective learning outcomes (short-term and long-term test performance), as well as self-perceived knowledge, subject confidence and enjoyment.

Methods

We conducted a randomized controlled trial within the Medical School at the University of Edinburgh, following ethics approval. Written consent was obtained from all participants. All information was handled in an anonymous and confidential manner.

All medical students from the first three years of the curriculum, or who were intercalating (undertaking an extra year of study to obtain a BSc), were invited to participate (approximately 700 students). We used oral and dental anatomy as our subject material in the study since this is not covered in the Edinburgh MBChB curriculum. Anatomy teaching at the University of Edinburgh does not normally utilize the ARS approach.

All respondents initially undertook an online 20-item multiple-choice test (referred to subsequently as the “pre-lecture test”). Students were asked to complete this as a “closed book” test, without reference to any knowledge source. Students were divided, at the median of pre-lecture test scores, into high and low score groups. These were used to perform stratified randomization of participants to a lecture either taught with an ARS or taught without an ARS. Randomization was performed using random number generation and students were not aware of their allocation in advance of the lecture.

The lectures took place on consecutive days, approximately one week after the pre-lecture test. They were given by the same lecturer using an identical format

and covering identical content. They differed only with regard to the inclusion of ARS questions, which were followed immediately by the lecturer briefly discussing the answers, with explanations tailored to the spread of answers received. Ten ARS questions were interspersed throughout the lecture. Questions were mostly multiple-choice and used the Interwrite Cricket ARS (Turning Technologies, Youngstown, OH). Individual handset responses were linked to answers in the pre-lecture/post-lecture tests.

Immediately following the lecture, students completed a paper-based 20-item multiple-choice test (hereafter the “immediate post-lecture test”). Students also completed a survey which collected demographic data and assessed student perceptions of their learning experience using five-item Likert scales. Finally, in order to test long-term knowledge retention, a further online “closed book” 20-item multiple-choice test was administered ten weeks after the lecture (hereafter the “final post-lecture test”).

Multiple-choice exam questions in the three test papers were mapped against the lecture’s learning outcomes to ensure a balanced coverage of topics. Neither post-lecture test contained questions posed to the ARS group during the lecture. This ensured that students were not simply recalling the wording of a previously encountered question.

A power calculation was performed using data from the ‘immediate post-test’ score in a similar trial [19]. Mean score in that study was 7.65 with standard deviation (SD) 1.16 in students taught without clickers. We estimated that a total

sample size of 98 would be required to detect a 10% improvement in score with power 0.9 and alpha 0.05.

Descriptive statistics were produced as appropriate to data type and distribution. Parametric tests were used only following acceptance assessment of data approximation to a normal distribution, and sufficient numbers. Analysis of test results was by comparison of groups, but was additionally explored using individually paired differences. Unless otherwise stated, correlation was tested using the Pearson method, which assumes that at least one variable is normally distributed. For all analyses, a p-value of <0.05 was regarded as significant. Statistical analysis was performed using Microsoft Office Excel 2003, Service Pack 3 (Microsoft Corp., Redmond, WA) and R-package software version 3.0.2 (The R Foundation for Statistical Computing, Vienna, Austria).

Results

Eighty of the approximately 700 invited students completed the pre-lecture test and were randomly assigned to lecture groups. A total of 70 students attended the lectures. No reasons for non-attendance were available for analysis. Thirty-nine students attended the lecture taught with an ARS (“ARS students”), and 31 attended the lecture given without an ARS (“non-ARS students”). Participant progress through the phases of the trial is shown in Figure 1.

Figure 1: CONSORT flow diagram (uploaded as separate file).

Participant characteristics are shown in Table 1 and did not vary between groups. All were undertaking medicine as their first undergraduate degree except for one ARS-student who had a previous undergraduate degree in dentistry. The control lecture lasted 47 minutes and the intervention lecture lasted 53 minutes. The multiple-choice tests had acceptable reliability with Cronbach alpha values of 0.68 for the immediate post-lecture test and 0.60 for the final post-lecture test.

Post-lecture survey

Post-lecture survey results were based on 68 students for some questions and 67 for others as 3 students failed to correctly complete the survey. There was no difference in levels of self-perceived knowledge between groups: mean 5-point score was 2.80 (SD 0.66) in ARS-students, versus 2.69 (SD 0.85) in non-ARS students ($p = 0.57$, 2-sample t-test).

When asked for their level of agreement with the statement “I now feel confident with the topic of oral anatomy”, 16/39 (41%) of ARS-students agreed or strongly agreed, versus 10/28 (36%) of non-ARS students ($p = 0.62$, χ^2 test)

When faced with the statement “I enjoyed this lecture on oral anatomy”, 37/39 (95%) of ARS-students agreed or strongly agreed, versus 24/28 (86%) of non-ARS students ($p = 0.25$, Fisher’s exact test).

When the 39 ARS-students were asked how the ARS had affected their learning experience, 30 stated “positively” and 7 “very positively” (95% in combination). No student felt it had negatively affected their experience.

Immediate post-lecture test performance

Immediate post-lecture test performance was significantly superior in the ARS group (score 17.3/20 versus 15.6/20, $p = 0.01$, Table 2). Test performance was superior in the ARS group in both genders and across all year groups, although in sub-group analysis this only reached statistical significance in males and in Year 2 students. These results were unaltered by using a paired differences analysis, comparing differences in individual scores between the pre-lecture and immediate post-lecture tests.

Final post-lecture test performance

The final post-lecture test was completed (online, ten weeks following the lecture) by 65/70 (93%) of students.

When performance was directly compared between groups, there was no difference, including in sub-group analysis (Table 2). A paired analysis comparing

differences in individual scores between the final post-lecture test and the pre-lecture test also showed no significant difference between groups (mean improvement 1.5 in the ARS group, versus 0.7 in the non-ARS group, $p = 0.31$). Final test score in the study population as a whole (mean 10.4/17) had significantly improved when compared to the baseline pre-lecture test (9.2/17), $p = 0.02$.

In the group taught with the ARS, there was no correlation between in-lecture ARS question performance and either immediate post-lecture scores (correlation 0.23, $p = 0.16$) or final post-lecture scores (correlation 0.11, $p = 0.52$).

Discussion

Summary

Our findings suggest that the introduction of ARS into anatomy education had no impact on students' self-perceived post-lecture subject knowledge, confidence or enjoyment, and had no demonstrable impact on long-term student performance. Several randomized controlled studies have investigated the use of ARS in other educational settings, and have not consistently demonstrated improvements in immediate or long-term knowledge tests [15]. Our observation of a modest improvement in performance immediately post-lecture , with its subsequent disappearance over an intervening ten-week period is in keeping with a study from Tregonning and colleagues who found that although ARS improved immediate knowledge test scores, this effect was not seen 5 weeks later [20]. A systematic study by Nelson et al concluded that a tendency to produce results strongly in favour of ARS is associated with non-randomized study designs [15].

Limitations

It should be noted that the present study was based on a single lecture intervention, allowing for a controlled experimental environment. Students had little time to incorporate the ARS approach into their expectations and learning style. It is conceivable that an intervention consisting of more than one lecture could have resulted in improvements in long-term student performance.

The failure of this study to recruit the numbers suggested by the pre-trial power calculation is a further limitation. Participation was voluntary and attendance at an evening lecture unrelated to the core MBChB curriculum may not have been perceived as valuable by non-participants. There was no identifiable reason for non-attendance from randomised participants, or for why this varied between groups (8/39 vs 2/41). The control lecture preceded the ARS-assisted lecture, and participants were blinded to their assignment. Despite these difficulties, our study was still able to demonstrate a significant difference in immediate post-lecture test results. Low numbers may have contributed to the failure to demonstrate a significant difference at ten weeks, and also rule out any firm conclusions on subgroup analyses. All conclusions in this study were unaltered by analysis using non-parametric tests and by using a paired-data approach.

Interpretation

Despite the lack of significant improvement in performance, the vast majority (95%) of students who received ARS teaching felt that its use positively affected their learning experience. Again, this is in keeping with previous findings [21]. It remains unexplained, therefore, why such a positive reported response from students was not reflected in improved perceptions of confidence or knowledge after the introduction of ARS. One parsimonious explanation is that ARS technologies were seen by students as an intervention that promoted 'intentional engagement' with their learning [22], marking a major change from their normal teaching methods and leading to a perception of positive change irrespective of whether the change actually impacted on their learning.

Despite the resource implications and training needs created by ARS[23, 24], its potential advantages remain attractive. ARS provide interactivity within the teaching setting, which would appear important, considering that in the long term large group teaching is not particularly effective in delivering student learning [25]. Knowledge retention following conventional teaching often decays at an undesirable rate [25, 26]. Post-teaching testing is a valuable way to improve retention of knowledge as it has been shown to slow decay of knowledge[27], and ARS may yet be shown to be a useful mechanism for this. Furthermore, the use of technologies such as ARS within teaching sessions could create what are known as a “learning landmark” exercises [28]. These are vivid experiences which are memorable in themselves and which then provide access to the educational content associated with that memory. Other examples include the use of virtual reality software in teaching, or body painting within an anatomy class [28, 29].

Conclusions

In conclusion, our study found that ARS use in a single lecture improved objective short-term anatomical knowledge, but did not improve long-term knowledge retention, when integrated into a traditional one-hour lecture format. Moreover, although students perceived ARS use to improve their learning experience, it did not improve self-perceived knowledge, nor subjective subject confidence. However, ARSs have good theoretical potential to improve the effectiveness of anatomy teaching and given the encouraging results of short-term knowledge

improvement we recommend further investigation of their use, including deployment in a more longitudinal fashion.

Acknowledgements

The authors would like to thank academic and technical colleagues at the University of Edinburgh for helpful discussions regarding study design and data interpretation, and for assisting with setting up the technology required for the ARS lectures.

References

1. Draper SW, Brown MI. Increasing interactivity in lectures using an electronic voting system. *J Comput Assist Learn.* 2004;20 :81–94.
2. Doucet M, Vrins A, Harvey D. 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.* 2009;31: e570–9.
3. Steinert Y, Snell LS. Interactive lecturing: strategies for increasing participation in large group presentations. *Med Teach.* 1999;21: 37–42.
4. Nayak L, Erinjeri JP. Audience response systems in medical student education benefit learners and presenters. *Acad Radiol.* 2008;15: 383–9.
5. Guse DM, Zobitz PM. Validation of the audience response system. *Br J Educ Technol.* 2011 Nov 1;42: 985–91.
6. Chaudhry MA. Assessment of Microbiology Students' Progress With an Audience Response System. *J Microbiol Biol Educ.* 2011;12: 200–1.
7. Heaslip G, Donovan P, Cullen JG. Student response systems and learner engagement in large classes. *Act Learn High Educ.* 2014;15: 11–24.
8. Abate LE, Gomes A, Linton A. Engaging students in active learning: Use of a blog and audience response system. *Med Ref Serv Q.* 2011;30: 12–8.
9. Turban JW. Students Prefer Audience Response System for Lecture Evaluation. *Int J Emerg Technol Learn IJET [Internet].* 2011;6. Available from: <http://online-journals.org/i-jet/article/view/1743>

10. Preszler RW, Dawe A, Shuster CB, Shuster M. Assessment of the Effects of Student Response Systems on Student Learning and Attitudes over a Broad Range of Biology Courses. *CBE-Life Sci Educ.* 2007;6: 29–41.
11. Caldwell JE. Clickers in the Large Classroom: Current Research and Best-Practice Tips. *CBE-Life Sci Educ.* 2007;6: 9–20.
12. Crouch CH, Mazur E. Peer Instruction: Ten years of experience and results. *Am J Phys.* 2001;69: 970.
13. Miller RG, Ashar BH, Getz KJ. Evaluation of an audience response system for the continuing education of health professionals. *J Contin Educ Health Prof.* 2003;23: 109–15.
14. Kirkwood A, Price L. Technology-enhanced learning and teaching in higher education: what is ‘enhanced’ and how do we know? A critical literature review. *Learn Media Technol.* 2014;39: 6–36.
15. Nelson C, Hartling L, Campbell S, Oswald AE. The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME Guide No. 21. *Med Teach.* 2012;34: e386–405.
16. Mastoridis S, Klaidis S. Coming soon to a lecture theatre near you: the ‘clicker’. *Clinical Teacher.* 2010;7: 97–101.
17. Alexander CJ, Crescini WM, Juskevitch JE, Lachman N, Pawlina W. Assessing the integration of audience response system technology in teaching of anatomical sciences. *Anat Sci Educ.* 2009;2: 160–6.
18. Border S, Walker F, Smith C. Are zappers a way of improving anatomy education? *Clin Anat.* 2009 Apr 1;22: 410–6.

19. 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.
20. Tregonning AM, Doherty DA, Hornbuckle J, Dickinson JE. The audience response system and knowledge gain: A prospective study. *Med Teach*. 2012;34: e269–74.
21. Kay RH, LeSage A. Examining the benefits and challenges of using audience response systems: A review of the literature. *Comput Educ*. 2009;53: 819–27.
22. Graffam B. Active learning in medical education: Strategies for beginning implementation. *Med Teach*. 2007;29: 38–42.
23. Halloran L. A comparison of two methods of teaching. Computer managed instruction and keypad questions versus traditional classroom lecture. *Comput Nurs*. 1995;13: 285–8.
24. Cain J, Robinson E. A primer on audience response systems: current applications and future considerations. *Am J Pharm Educ*. 2008;72: 77–77.
25. Bligh DA. 1998. *What's the use of lectures?* Exeter, Intellect.
26. Sawdon MA. Improving knowledge retention using KEEpad. *Med Educ*. 2009;43: 487.
27. Karpicke JD, Roediger HL III. The critical importance of retrieval for learning. *Science*. 2008;319: 966–968.
28. Finn GM, McLachlan JC. A qualitative study of students' perceptions of body painting. *Anat Sci Educ*. 2010;3: 33-38.
29. Finn GM. Using Body Painting and Other Art-Based Approaches to Teach Anatomy. Chan LK, Pawlina W, editors. *Teaching Anatomy: A Practical Guide*. 2015. pp. 155-164.

Tables

Table 1: Overview of participant characteristics

Characteristics	Taught with ARS n (%)	Taught without ARS n (%*)
Gender		
Females	21 (54%)	15 (52%)
Males	18 (46%)	14 (48%)
Year group		
Year 1	13 (33%)	14 (48%)
Year 2	17 (44%)	12 (41%)
Year 3	7 (18%)	2 (7%)
Intercalated¶	2 (5%)	1 (3%)
Pre-lecture test score (maximum score of 20): mean (SD)	9.0 (1.9)	9.5 (2.8)

* Percentages in this column are for 29/31 participants in this group due to missing data (two students failed to complete their post-lecture survey).

¶ Intercalated students are medical undergraduates completing an additional one year period of study which leads to a B.Sc degree and is undertaken after Year 2 of their medical degree.

Table 2: Immediate and final post-lecture test scores

Participant characteristics	All students (mean score, SD)	Taught with ARS (mean score, SD)	Taught without ARS (mean score, SD)	<i>p</i> -value (ARS versus non-ARS group)
Immediate post-lecture test scores (maximum score of 20)				
Male	16.6 (3.4)	17.7 (2.5)	15.1 (3.7)	0.02
Female	16.6 (2.2)	17.0 (2.2)	16.1 (2.1)	0.25
Year 1	15.8 (2.4)	16.0 (2.3)	15.6 (2.5)	0.65
Year 2	17.2 (2.3)	17.9 (2.1)	16.2 (2.2)	0.04
Year 3/ intercalated	17.2 (4.2)	18.2 (2.1)	14.0 (7.8)	0.39*
All group participants	16.6 (2.8)	17.3 (2.3)	15.6 (3.0)	0.01
Final post-lecture test scores (maximum score of 20)				
Male	10.1 (3.6)	9.9 (3.1)	10.6 (4.4)	0.61
Female	10.7 (2.9)	11.1 (3.2)	10.1 (2.3)	0.33
Year 1	8.9 (3.0)	8.7 (3.1)	9.1 (3.0)	0.75
Year 2	11.0 (2.8)	11.3 (3.0)	10.6 (2.5)	0.54
Year 3/ intercalated	12.5 (3.4)	11.8 (2.6)	16.0 (5.7)	0.18*
All group participants	10.4 (3.2)	10.5 (3.2)	10.1 (3.4)	0.61

Significance testing was carried out using 2-sample t-tests, except in the rows indicated with a *, where the Wilcoxon rank sum method with continuity correction was used owing to small numbers.