



<https://www.aims sciences.org/journal/A0000-0006>

Research article

Embedding opportunities for participation and feedback in large mathematics lectures via audience response systems

Christopher C. Tisdell^{1,2,*}

¹ School of Mathematics and Statistics, The University of New South Wales, Sydney NSW 2052, Australia

² Institute for Teaching and Learning Innovation (ITaLI), University of Queensland, Brisbane Qld 4072, Australia

* **Correspondence:** cct@unsw.edu.au; Tel: +61-2-93856792

Academic Editor: William Guo

Abstract: The purpose of this work is to interpret the experiences of students when audience response systems (ARS) were implemented as a strategy for teaching large mathematics lecture groups at university. Our paper makes several contributions to the literature. Firstly, we furnish a basic model of how ARS can form a teaching and learning strategy. Secondly, we examine the impact of this strategy on student attitudes of their experiences, focusing on the ability of ARS to: assess understanding; identify strengths and weaknesses; furnish feedback; support learning; and to encourage participation. Our findings support the position that there is a place for ARS as part of a strategy for teaching and learning mathematics in large groups.

Keywords: mathematics education, audience response systems, feedback, large group teaching, open educational resources, digital education

1. Introduction, Background and Motivation

Lectures (face-to-face, blended and online forms due to COVID-19) have continued within many universities despite much discourse regarding their limitations. For instance, traditional, one-way instructional approaches to lecturing, including within large, face-to-face groups, have come under criticism for decades due to a focus on content delivery and the risk of passive learner behaviour [7]. The continuation of lectures in the large group setting may be partly due to the perceived economic efficiencies of their “one lecturer to many students” operating model. However, determining how to balance delivery and active listening; with creating opportunities for activities, assessment and feedback continues to pose a significant challenge within large lecture groups.

Audience response systems (ARS) provide an alternative perspective to orthodox, one-way lecturing. While there is no universally accepted definition of ARS, the idea is to repurpose face-to-face lecture time for assessment, discussion and feedback. By embedding opportunities for more activities in lectures, the concept of ARS can be pedagogically anchored in the theory of active learning [8], where students: are engaged in meaningful tasks; and are thinking about what they are doing, or reflecting on what they have done.

The aforementioned dimensions of ARS can be connected with the wider theme of flipped classrooms [4]. Although the interest in, and practice of, flipped mathematics classrooms is increasing, the notion of flipping remains under-researched and sub-theorized within this environment. There are few research studies involving flipping mathematics classes at the tertiary level [34]. Indeed, student perceptions of the flipped classroom and its associated learning outcomes remain largely unexplored [10]. In addition, there is little widespread evidence regarding what kinds of flipping works best and for whom.

Furthermore, the findings of initial research to probe students' perceptions of flipped mathematics classes have been mixed. For example, Johnston [24] found that while students displayed some positive responses to certain elements of a flipped classroom, the overall student satisfaction for the course under consideration was not higher than those of previous years, when taught using conventional lectures. Novak et al [36] revealed that only 10% of the students in their study believed that flipping helped them to learn better in general; whereas Jungic et al [25] found that many students perceived that their flipped class helped them to learn better than the traditional format. While Love et al [30], Murphy et al [33] and Cronhjort et al [17] all reported increased levels of enjoyment or confidence for students in a flipped mathematics classroom; Petrillo [38] discovered up to half of students in some classes did not perceive their experience in a flipped classroom as effective.

The results of research studies on the effect of flipping mathematics classrooms on learning are varied. For instance, Petrillo [38] saw a decrease in the failure rate of students in flipped classrooms; and Maciejewski [31] and Cronhjort et al [17] both reported learning gains of 8% and 13%, respectively, where students in flipped classrooms outperformed those in the traditional lecture environment. Alternatively, Johnston [24] and Wasserman et al [51] found no significant differences in course or examination results between students in a flipped classroom and students in traditional courses. Naccarato & Karakok [34] reports Bagley [3] identifying students in a flipped section being "outperformed" by those in a traditional model on procedural questions in the final exam.

While there have been some studies undertaken that involve demographics, including gender difference, in flipped mathematics classrooms [10], current research just scratches the surface of the question - what kinds of flipping are successful, and for whom?

There have been various approaches to ARS, both with and without digital technology. Perhaps the most familiar method involving ARS is for the teacher to pose a question to the class and the students respond by simply raising their hands. Decades ago, Bligh [7] discussed various "low-tech" forms of ARS, including the use of coloured cards or cubes that students could hold up to "vote" in response to a question. While these appear to be straightforward, time-honoured and easily implementable initiatives, they face challenges regarding anonymity and in interpreting the response data at scale.

In more recent times, the area of ARS has drawn on the field of digital education, through "innovative use of digital tools and technologies during teaching and learning" [49], including clickers and mobile devices. Banks [5] and Duncan [19] provide a rich tapestry of research on ARS in the form

of electronic clickers / handsets and associated software, where students can respond by selecting the appropriate button on their clicker that is associated with an answer. In addition, mobile devices have been employed as ARS within mathematics education in schools, see for example, Larkin & Calder [28] and the associated journal special issue in which it appears.

However, current understanding regarding the impact of digital ARS in mathematics classes remains mixed, which the following two examples illustrate. The results obtained by King & Robinson [26] show that ARS use: “has yet to make any difference to student performance, as measured by mean grades obtained on a Mathematics module”; and “had a negligible effect on student attention and retention rates.” On the other hand, studies from Dunn et al [20]; and Dunn et al [21] suggest that the use of ARS can increase student engagement within large statistics classes.

When the area of ARS is viewed through the lens of digital education, much of the implementation and research has involved software of a commercial nature, including subscription-based models (to name just two: Voteapedia in [20]; eduVote in [40]). The preceding situation naturally raises economic questions regarding value-for-money and sustainability. There are also equity and accessibility challenges for those who use it - will students still be able to access the software if the funding for the software is not continued? Within this context of software considerations for ARS, perspectives involving open educational resources (OER) have mostly remained sheltered from discussions. OER may be defined as “digitised materials offered freely and openly for educators, students, and self-learners to use and reuse for teaching, learning, and research” [37].

To summarize our discussion from above, there is an opportunity and a need to explore gaps regarding current understanding of ARS when viewed through the lenses of digital education and OER in large mathematics classrooms. The purpose of our present study is to interpret the experiences of students when ARS was implemented as a strategy for teaching large mathematics lecture groups at university, drawing on elements of OER and digital education.

To explore the aforementioned opportunities, we draw on case study research. Case study research has become a popular approach in the social sciences [18, p.114] and involves “an empirical investigation of a particular contemporary phenomenon within its real-life context using multiple sources of evidence” [39]. The power of case study research lies in its ability “to enable the research to intensively investigate the case in depth, to probe, drill down and get at its complexity” [18, p.114].

Creswell [16] and Yin [52] emphasize that a suitable strategy to steer case study research is asking questions of the type “how” and “why” to better understand a phenomenon. In addition, research questions asking “what” are appropriate if the objective is to explore and illuminate an event [52]. With these elements in mind, our work is guided by the following research questions:

- RQ1: In what way can ARS form part of a teaching and learning strategy in large mathematics classrooms?
- RQ2: How can such a teaching strategy impact student attitudes regarding its ability to: assess understanding; identify strengths and weaknesses; furnish feedback; support learning; and encourage participation?
- RQ3: What does the experience of students concerning an implementation of ARS suggest about its potential for teaching and learning mathematics within large groups?

We respond to RQ1 through reflection on, and discussion of, the design and delivery of our intervention involving ARS, and identify key components. We address RQ2 and RQ3 through a quasi-

experiment that employs survey techniques and a resulting analysis.

We organise our work in the following way: In Section 2 we present our research design and methods. This includes discussing the intervention in more detail; establishing and defending our methodological position; and examining our groups of interest. Section 3 contains our results, analysis and discussion. In Section 4 we examine the limitations with our study by considering counterfactual perspectives. Finally, we present our conclusions and recommendations for further work in Section 5.

2. Research Design, Methods and Approach

2.1. Position on Methodology and Quasi-Experiments

Our methodological position throughout this work is based on the presumption that all experiments are biased, but some of them are still useful. Thus, our point of view is influenced by rationalism [32], post-positivism [48] and critical realism [2].

To evaluate the impact of our intervention involving ARS, digital education and OER, we designed and delivered a quasi-experiment. Quasi-experiments share the aims of all other experiments, to wit, “to test descriptive hypotheses about manipulable causes, as well as many structural details, such as the frequent presence of control groups and pretest measures, to support a counterfactual inference about what would have happened in the absence of treatment” [42].

The inclusion of a control group that forms part of quasi-experimental evaluation design is favoured over non-experimental approaches, such as before-and-after design, for instance, due to the susceptibility of a before-and-after design to internal validity threats [39].

Quasi-experiments provide an appropriate alternative to classical experiments in cases when randomization is impractical or unethical [12]. Due to randomization being absent, this approach “provides a limited counterfactual which can infer limited causation” [12]. For instance, it does not control for selection bias.

A key instrument for our quasi-experiment is the employment of surveys. In Wang [50, p.36], we observe reference to the term *survey* as “an instrument to collect data that describes one or more characteristics of a specific population.” Survey research is suitable for use in education due to its ability to gather information about population groups to “learn about their characteristics, opinions, attitudes, or previous experiences” [50, p.128]. It is through this collection of data and its analysis that this method of research provides insight into the attitudes, thoughts and opinions of populations. Survey methods have grown in popularity over the past few decades to form an important, accepted, cost-effective and time-efficient way of doing research within the social sciences [6].

2.2. Domain of Inquiry and Groups of Interest

Let us discuss the environment within which our intervention took place. Our mathematics course MATH1131 is usually undertaken by students within their first year of studies at The University of New South Wales - a large institution with in-excess of 50,000 students. MATH1131 is a compulsory course for those pursuing such fields as engineering, chemistry, mathematics education and physics. Thus our particular course can be aligned with the domain of “service teaching” as the students therein were not necessarily specializing in mathematics.

MATH1131 is a relatively large course, with in-excess of 1,600 students. Its syllabus consists of calculus and algebra. Lectures were conducted face-to-face and lecture theatre capacity constraints

necessitated that there were multiple parallel lecture groups operating. A student would experience two lectures per week on calculus; and two lectures per week on algebra. Students were unable to take their two hours of calculus within one lecture group and then switch to a different group to take their two hours of algebra. In this sense, there was continuity in the student cohort between calculus and algebra.

The groups of interest in this research study are identified in Table 1.

Table 1. Groups of Interest

Group	Details
Target:	Undergraduate students in large mathematics classes
Sample:	Students in Lecture Group 1 of MATH1131 during the algebra lectures where the intervention took place
Comparison:	Students in Lecture Group 1 of MATH1131 during the calculus lectures where no intervention took place

In our research design we have formed Sample and Control groups in such a way that the sample group of students experience the intervention each week within the two-hourly algebra lectures of the course (this forms the Sample Group); but they do not experience the intervention within the two-hourly calculus lectures of the same course (this forms the Control Group). When we thus ask them common questions of their experiences, this approach forms a natural opportunity to compare their attitudes towards each situation because each student has experienced the intervention in the Algebra lectures, but has not experienced the intervention in their Calculus lectures. This partially controls for diversity in the student population.

Randomization of students for this study was impractical due to timetable constraints. Students chose to enrol in a particular lecture stream based on their commitments, preferences and timetable. Randomly distributing students into a range of groups would have risked timetable clashes with their other courses. Thus, a random assignment of students was not possible in this case, meaning that it was well-suited to the quasi-experimental approach.

Students were unaware that the intervention was going to take place when they chose their lecture group to enrol in, reducing the threat of self-selection.

2.3. The Intervention in More Detail

Let us respond to RQ1 through reflection and discussion of how we designed and implemented ARS to form our teaching strategy. We form themes of ARS, OER and digital education, where each are mapped against their components in Table 2, forming a basic framework.

Table 2. Themes and Relevant Components of Audience Response Systems

Theme	Relevant Components
ARS:	Lecture (re)design and (re)delivery Embedding of discussion, assessment and feedback
Digital Education:	Use of mobile devices (laptops, phones, tablets) Creation of YouTube videos
Open Educational Resources:	Use of Google Forms

Because we were repurposing our lecture time, our redesign involved making decisions regarding what to do with previous lecture material. Which examples were we to keep and discuss in the shorter timeframe? Which to embed into the assessment activities; and which to remove? We employed a “comparative judgement” approach to assist with these decisions, ranking the examples from what we believed were most important (i.e., keep and discuss or test), to least important (remove and repurpose). The examples that we removed from lectures were repurposed into YouTube videos, aligning with the theme of digital education. Students could access these videos after (or before) class as a complementary learning resource. This element thus shares some characteristics with the flipped classroom in mathematics [34], [10], [24], [51], although we did not mandate the students to watch the videos.

Our redesign led to a trimmed timeframe for **delivery** within lectures. This signified and embodied a shift in thinking of the role of lectures from “delivery of content” to a more varied experience for the students where feedback, interaction and content could all play a role.

In lectures we formatively **assessed** the class after the delivery was paused, aligning with the theme of ARS, and to see if students had grasped a topic [29]. Students were given a shortened URL that they could type into their mobile devices (phones, laptops, tablets) which would take them to a Google Form. Google Forms enable educators to create an online quiz via a range of questions, from multiple choice to dropdowns, to a linear scale [22]. Google Forms strongly aligns with the theme of OER due to its free and open nature. For instance, no subscription costs are involved, no login is required by students to access it, and the quizzes work on a wide variety of devices and browsers.

For the design of our quizzes, we settled on the multiple choice question format for simplicity, and formatted the questions with large font so that those students who were using small screen devices could more easily view the questions in detail. Since Google Forms does not support LaTeX, we opted to use Codecogs [11] to mathematically typeset the questions and produce output in pictorial format. Some example questions on complex numbers are included in Figures 1-3.

- A geometric relationship between z and its conjugate \bar{z} involves:
- A reflection in the real axis
 - A reflection in the imaginary axis
 - A rotation about the origin of one half of a revolution
 - None of the above.

Figure 1. Sample Question from Quiz: Q1ii)

If $z = \frac{1}{i}$ then its conjugate \bar{z} is:

- i
- $-i$
- 1
- -1

Figure 2. Sample Question from Quiz: Q2i)

Complete the identity: $\frac{1}{2}(z + \bar{z}) =$

- a) $Re(z)$
- b) $Im(z)$
- c) z^2
- d) $1/z$

Figure 3. Sample Question from Quiz: Q2ii)

Students were encouraged to discuss the questions in pairs before responding. This was to foster peer-to-peer interactivity and to accommodate students who may not have had access to a personal mobile device in the lecture.

Once the students had submitted their anonymous responses to the quiz, the summarized results could be displayed on the projector screen via the graphs of Google Forms for all to see. Some examples of responses to those questions in Figures 1-3 are contained in Figures 4-6. This is what would be seen by all of the class, including the lecturer. (We note that the 100 responses would comprise groups of students working together to submit one set of answers between them.)

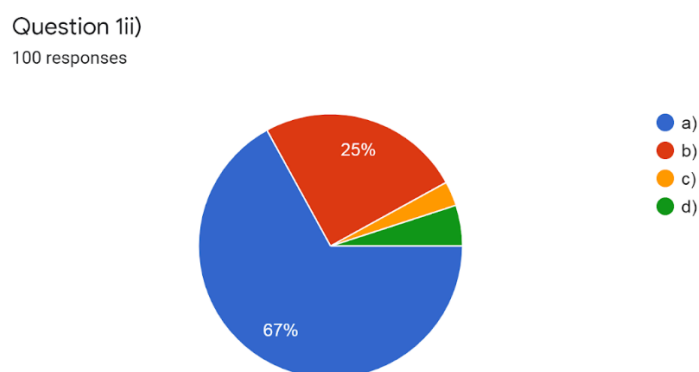


Figure 4. Class Responses to Question 1ii) of Quiz

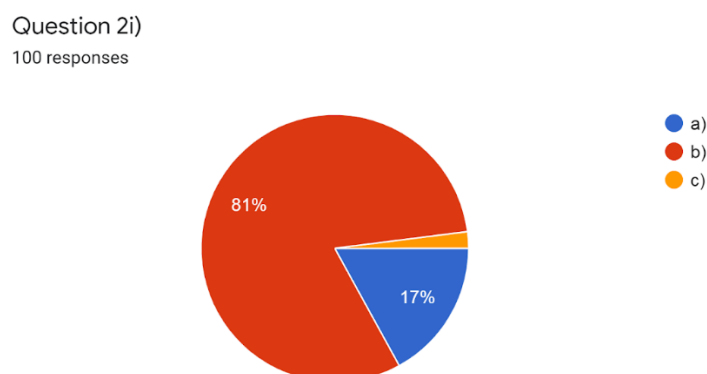


Figure 5. Class Responses to Question 2i) of Quiz

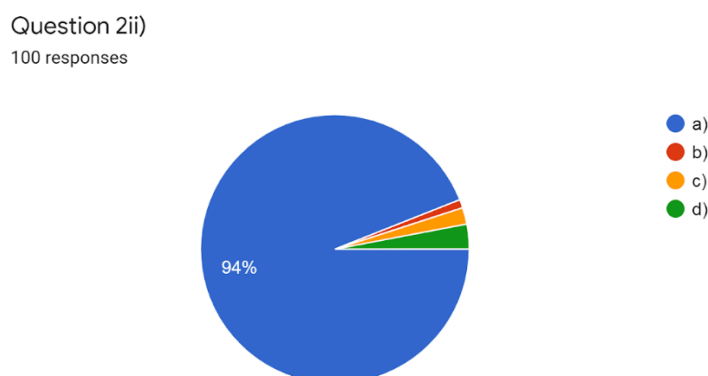


Figure 6. Class Responses to Question 2ii) of Quiz

The graphs enabled the lecturer to provide **feedback** on the results, aligning the theme of ARS. We could identify where students had correctly or incorrectly responded, and it also gave each student some idea about how their individual responses (or responses from pairs) compared with the overall class responses.

A **discussion** involving the analysis of the results from the quiz then could take place, and if there were common errors or misunderstandings identified, then those items could be revisited and emphasis placed on how students might **improve**. For example, it is clear that only a minor discussion would be required based on the data from Figure 6, however the data from Figures 4 and 5 would warrant some additional discourse with more than 25% of incorrect responses. Both of these actions align with our theme of ARS.

Although the length of time involved for completing an ARS cycle did vary depending on the amount of discussion taking place, we estimate time bounds for each ARS intervention at 10-15 minutes. This can be roughly broken down into: 5-7 minutes for students to discuss and complete the quiz; and 5-7 minutes to analyse and discuss the responses. Usually three or four questions were asked per quiz and they covered a variety of concepts to foster student thought regarding a variety of perspectives. For example: geometry or diagrammatic thinking (for example, see Figure 1); computation (for example, see Figure 2); and other aspects (for example, see Figure 3). These roughly align with the “Quick Check” style of questions outlined in [29].

A basic framework of how we implemented our intervention in lectures viewed through the lens of sequencing and technology is summarized in Table 3.

Table 3. Summary of Intervention: Framework for Delivery

Activity	Technology
Deliver material at start of lecture	None necessarily required
Assess material via short, formative quiz	Accessed via Google Forms / responses via m-devices on wifi or network
Feedback to class on the results of quiz	Discuss results via graphs from Google Forms
Discussion and thoughts on how to improve (if needed)	None necessarily required

3. Instruments, Analysis and Results

3.1 Evaluation Overview

We summarize our evaluation overview in Table 4, unpacking some of its elements in the subsections that follow.

Table 4. Evaluation Overview

Evaluation Approach	Timing/Sample/Analysis	Evaluation Focus
Attitude Data 1: (Sample Group)	Post intervention. Attitudinal data collected from bespoke survey of 348 students within component of intervention (algebra lectures). Six-point Likert scale employed; mean values calculated, including 95% confidence intervals. Comments collected and coded.	Impact on students' attitudes towards their learning experience.
Attitude Data 2: (Sample Group)	3 weeks after Survey 1. Attitudinal data collected from survey of ~180 students within component of intervention (algebra lectures). This is a subset of the 348 students from previous survey. Six-point Likert scale employed; mean values calculated, including 95% confidence intervals. Comments collected and coded. Sample Group and Control Group compared via statistical tests.	Impact on students' attitudes towards their learning experience.
Attitude Data 3: (Control Group)	3 weeks after Survey 1. Attitudinal data collected from survey of 102 students within component where no intervention took place (calculus lectures). This is a subset of the 348 students from previous survey. Six-point Likert scale employed; mean values calculated, including 95% confidence intervals. Sample Group and Control Group compared via statistical tests.	Impact on students' attitudes towards their learning experience.

3.2 Design of Survey Questions

To better understand student perceptions of our intervention and their attitudes towards learning, we invited students to participate in some surveys.

The statements presented to students are contained in Table 5. The set of statements A-F therein was shared with the Sample Group via Survey 1. The set of statements G-H was shared with both the Sample Group and the Control Group via Survey 2.

Table 5. Statements in Survey 1 (A-F) and Survey 2 (G-H)

Item	Statement
A	The quizzes provided a valuable opportunity to test my understanding of basic ideas
B	The quizzes helped to identify specific strengths and weaknesses of my understanding
C	It was valuable to have immediate feedback and discussion of the results
D	The quizzes encouraged and supported my learning
E	Overall, I was satisfied that these quizzes were a valuable learning tool
F	I would like to have these kinds of quizzes available to support my learning in future lectures
G	This lecturer provided feedback to help me learn
H	This lecturer encouraged student input and participation during classes

We decided to employ our own designed instrument in conjunction with an instrument designed by our institution. Statements A-F were drawn from the Higher Education Research and Development Society of Australasia (HERDSA) Fellowship literature [23], namely Criterion 2: Assessment encourages and supports learning. Statements G-H are standard items within our institutional teaching surveys.

An analysis of each statement A-H in Table 5 reveals a strong alignment with the nature of our research questions RQ2 and RQ3 in the Introduction. Thus, while we acknowledge the leading nature of some of these questions, we believe the statements take an appropriate form for probing RQ2 and RQ3.

3.3 Survey Data

In each survey, students were asked to respond to the statements in Table 5 by selecting either: Strongly Disagree; Disagree; Mildly Disagree; Mildly Agree; Agree; Strongly Agree. According to Allen and Seaman “there's really no wrong way to build a Likert scale” [1] and they advocate to use as wide a scale possible with an upper limit of seven choices. Our choice of 6 points in the Likert scale is suitable for several reasons. Firstly, it aligns with our institution's practice regarding survey scales. Students at our institution are used to making evaluations on this scale and this builds confidence and familiarity in their choices, with respondents more able to apply them in making finer stimulus distinctions. Secondly, employing an even number of choices avoids what is described as a “doctrine of the mean” [15], where there is a tendency for participants to opt for the mid-point. Thirdly, participants were not forced to make a choice regarding any of these questions. If they did not wish to answer then they could simply leave it blank.

The responses to the surveys are tabulated in Table 6 and Table 7. The analysis of collected quantitative data is presented in Table 8 and Table 9. In the analysis of the data, a response of: Strongly Disagree is assigned the value of 1; Disagree = 2; Mildly Disagree = 3; Mildly Agree = 4; Agree = 5; finally, through to Strongly Agree being assigned the value of 6.

Table 6. Responses of Sample Group to Survey 1 and 2

Statement	Strongly Disagree	Disagree	Mildly Disagree	Mildly Agree	Agree	Strongly Agree	n
A	3	1	1	16	149	178	348
B	3	2	7	44	157	135	348
C	3	0	2	18	115	210	348
D	3	2	1	46	165	131	348
E	3	2	1	24	163	155	348
F	3	0	1	15	136	193	348
G	1	4	2	24	66	81	178
H	0	1	1	4	44	131	181

From the mean scores summarized in Table 8 we can see that the levels of agreement ranged: between 5.17 (± 0.08) and 5.66 (± 0.10) on the 6-point Likert scale; and between 97% and 99% on the Overall Agreement scale. As these scores are above 5 (even at the extreme lower end of the confidence interval) and close to 100%, we interpret this as illustrating that overall student attitudes were positive

towards their experiences of our intervention of ARS as part of a teaching and learning strategy.

The two highest scores were in response to statement items C (5.51 ± 0.08) and H (5.66 ± 0.10), suggesting that students felt most positively about the value of the ARS regarding immediate feedback and discussion; and its ability to foster student input and participation during classes.

Table 7. Responses of Control Group to Survey 1 and 2

Statement	Strongly Disagree	Disagree	Mildly Disagree	Mildly Agree	Agree	Strongly Agree	n
G	2	10	17	37	26	10	102
H	5	11	22	37	19	9	103

Table 8. Analysis of Responses of Sample Group to Survey 1 and 2

Statement	Mean Score / 6	Confidence Interval 95%	% Overall Agree*	Standard Deviation of Mean	n
A	5.42	± 0.08	99	0.75	348
B	5.17	± 0.09	97	0.82	348
C	5.51	± 0.08	99	0.75	348
D	5.19	± 0.09	98	0.82	348
E	5.31	± 0.08	98	0.82	348
F	5.47	± 0.08	99	0.72	348
G	5.21	± 0.14	96	0.94	178
H	5.66	± 0.10	99	0.70	181

*Overall Agreement is defined as those responses of: Mildly Agree; Agree; or Strongly Agree.

The two lowest scores involved statement items B and D. We note that even at the lowest end of the confidence interval each of these scores still exceeds 5, suggesting a lower, but still solid level of positivity towards the elements of identification of strengths and weaknesses; and encouragement and support of learning when compared with responses to items C and H.

In Table 9 we see the summary of responses of the Control Group to Survey 2. A visual comparison between this and Table 8 strongly suggests that there are differences between responses from the two cohorts. We shall probe this further a little later.

Table 9. Analysis of Responses of Control Group to Survey 2

Statement	Mean Score / 6	Confidence Interval 95%	% Overall Agree*	Standard Deviation of Mean	n
G	4.03	± 0.23	72	1.18	102
H	3.79	± 0.24	63	1.25	104

*Overall Agreement is defined as those responses of: Mildly Agree; Agree; or Strongly Agree.

Not all feedback was positive, however, and some unfavourable responses were present as can be seen from Table 6 and Table 7.

3.4 Qualitative Comments

At the end of Survey 1, there was space for students in the Sample Group to provide some optional, free-text comments in case they wished to add to, or provide another dimension, to their feedback. Below, we present some qualitative perspectives on their experiences regarding this.

There were 72 unique comments from the Sample Group for Survey 1 which were directly relevant to our intervention. We coded these responses into three themes via use of NVivo 12: *Efficacy*; *Appreciation*; and *Constructive Suggestions*. Some comments were coded into multiple themes, for example “Keep them in the final parts of the lesson but otherwise very helpful” would fall into *Efficacy* and *Constructive Suggestions*. We summarize the number of comments corresponding to each theme in Table 10.

Table 10. Themed Comments from Sample Group in Survey 1

Theme	Number
Efficacy	32
Appreciation	29
Constructive Suggestions	16

A representative sample of comments regarding the theme of *Efficacy* includes:

- I really enjoyed the interactivity in the lecture with the quiz
- They are a good break
- Probably one of the best support tools I've seen in a lecture
- Great to break up the lecture and have immediate feedback
- they make the lesson interesting and interactive.

The comments coded into *Appreciation* mostly showed gratitude or a desire for the quizzes to continue, without going into further details. This included comments such as:

- thank you!
- i loved them
- Please make more of them!

The final theme of *Constructive Suggestions* contained comments regarding perceived limitations or how the experience might be improved. This included:

- Be more clear regarding the answers to the quiz
- Put one difficult question in
- I'd like to go through the answers more thoroughly
- Better to have more discussion so that we can help each other and know each other :)
- More time to do them and more time to explain the answers.

From the above perspectives, we can see that students' experiences were mostly of a positive nature, with some reasons for this identified as: interactivity; breaking up the lecture; feedback; and support.

There was feedback regarding perceived contours and limitations of the quiz, including the desire for more allocated time to complete and discuss the quiz; and the inclusion of more challenging questions.

3.5 Comparing the Sample Group with the Comparison Group

In Table 11 we compare the results from the common survey questions between the Sample Group and the Control Group where we have drawn on the data in Table 8 and Table 9.

Table 11. Significance Tests Between Sample Group and Control Group to Survey 2

Statement	Student's t-test $p < 0.05?$	Mann-Whitney U-test $p < 0.05?$	Effect Size (Cohen's d)
G	Yes	Yes	1.10
H	Yes	Yes	1.84

There has been healthy discourse on choice of pertinent comparison tests regarding ordinal data, especially for Likert scales used in surveys [1]. Traditionally, non-parametric tests (such as Mann-Whitney's U-test) have been assessed as being appropriate, however, when sufficiently large sample sizes are available, parametric tests (such as Student's t-test) have also been employed in the literature. Drawing on [43] we take the position that it is possible to employ both kinds of tests for our case and so have included both sets of findings for completeness. In this regard, we see from Table 11 that both tests show statistically significant differences between both questions with $p < 0.05$.

Acknowledging Cohen's position [13] that “The primary product of a research inquiry is one or more measures of effect size, not p values”, we draw on the advice of Glass (cited in [27]) regarding the importance of reporting results “in terms of measures of magnitude –not just, does a treatment affect people, but how much does it affect them.” Aligning with this position, Cohen's d-test is a well-known approach that measures “effect size” aimed at giving an indication of the size of the difference between two groups. From Table 11 we see the values of $d = 1.10$ and $d = 1.84$. Respectively, this suggests a large [13] and very large [41] effect of our intervention on the Sample Group compared against the Control Group with respect to the elements of: feedback to help learning; and encouragement of student input and discussion during classes.

4. Limitations

Let us discuss some limitations of this study by considering several threats to validity.

We argue that quasi-experiments serve as mere estimations to reality, and there is no need to ask *Is an experiment accurate?* For if “accurate” is to mean “entirely accurate” then the answer must be no. Instead, the question of interest is *Is the experiment illuminating?* Our perspective is informed by Box's position on scientific models [9] and guided by von Neumann's principle that “truth ... is much too complicated to allow anything but approximations” [35].

Additionally, we acknowledge two main threats to our work: the singular nature of the intervention; and non-randomization of participants. Our intervention took place during one semester. If we found consistent findings after conducting the intervention multiple times then this would strengthen and

support the case. We call for more research that considers randomization as part of the process. We also cannot absolutely rule out the possibility of selection bias having some influence during our intervention (for example, one semester's students may have been more enthusiastic towards ARS than other years). As we explained earlier, randomization was impractical due to timetable constraints.

Furthermore, a more powerful dataset may have been generated by combining student quiz performance and students survey responses. Due to the anonymous nature of quizzes and surveys, a link was not possible. This could be an interesting consideration in future studies.

We also cannot rule out effects on the results from those undertaking the teaching, and the subject-matter being taught. For example, we have not controlled for student's potential bias for (say) calculus over algebra, or for one particular lecturer over another. This presents an opportunity to explore these ideas in future work.

Herein we have examined what has worked locally “for us”. Aligning with the theory of case study research [52], a generalisation to larger populations based on a single case study has not been our aim. Nevertheless, our very selection of a case that involves ARS indicates that we are connecting it with a more comprehensive group of studies at universities, colleges and polytechnics; steering towards a collective understanding of their benefits and limitations within large mathematics classes.

5. Conclusion

Based on our findings, we can now respond to the three research questions posed in Section 1.

- RQ1: We put forth a specific model as a way of designing and implementing ARS as part of a teaching strategy for large groups. Key components were identified as: lecture redesign and redelivery to free up time; embedding of discussion, assessment and feedback; use of mobile devices; creation of YouTube videos; and use of Google Forms.
- RQ2: Our data and analysis suggest a strong and positive impact of our integrated teaching strategy on student attitudes towards: assessing understanding; identifying strengths and weaknesses; obtaining feedback; supporting learning; and encouraging participation. There were positive, statistically significant and large to very large effects on: feedback to help learning; and encouragement of student input and discussion during classes.
- RQ3: The experience of students with our integration of ARS supports the position that there is a place for this kind of approach for teaching and learning mathematics in large groups.

This work and its findings have aimed to “shine a light” in the aforementioned places and ways. We call on scholars and researchers to add more insight regarding ARS by conducting additional research on what works best and for whom. For instance, we see significant value in additional investigations involving: larger populations; a consideration of demographics; and effects (if any) on test performance. Can deeper, more time-heavy questions be asked [29] and what are their limitations? By engaging with these ideas scholars can open new and alternative understandings [46, 47], ordering and ways of working in our pedagogical landscape [44, 45] of ARS in large mathematics classes.

References

1. Allen, I.E., Seaman, C.A., *Likert scales and data analyses*. Quality Progress, 2007, **40**(7): p.

- 64-65.
2. Archer, M.S., Bhaskar, R., Collier, A., Lawson, T., Norrie, A. *Critical Realism: Essential Readings*. 2009, London, UK: Routledge.
 3. Bagley, S.F., *Improving student success in calculus: a comparison of four college calculus classes* [dissertation]. 2014, San Diego State University: San Diego, USA.
 4. Baker, J.W., *The 'classroom flip': using web course management tools to become the guide by the side*, in *Selected Papers from the 11th International Conference on College Teaching and Learning*. 2001, Florida Community College at Jacksonville: Jacksonville (FL), p. 9–17.
 5. Banks, D.A., *Audience Response Systems in Higher Education: Applications and Cases*. 2006, Hershey, PA, USA: Information Science Publishing.
 6. Berends, M., *Survey methods in educational research*, in *Handbook of Complementary Methods in Education Research*, J.L. Green, G. Camilli, P.B. Elmore, Ed. 2006, Lawrence Erlbaum Associates. p. 623-640. Retrieved from <http://psycnet.apa.org/record/2006-05382-038>
 7. Bligh, D.A., *What's the Use of Lectures?* 1972, Harmondsworth, UK: Penguin Books.
 8. Bonwell, C.C., Eison, J.A., *Active Learning: Creating Excitement in the Classroom*. 2005, San Francisco, USA: Jossey-Bass.
 9. Box, G.E., *Robustness in the Strategy of Scientific Model Building*. 1979, Ft. Belvoir: Defense Technical Information Center. Retrieved from <http://www.dtic.mil/docs/citations/ADA070213>
 10. Chen, S., Yang, S.J., Hsiao, C., *Exploring student perceptions, learning outcome and gender differences in a flipped mathematics course*. *British Journal of Educational Technology*, 2016, **47**(6): p. 1096-1112. doi:10.1111/bjet.12278
 11. Codecogs. Retrieved from <https://www.codecogs.com/latex/eqneditor.php?latex=D>
 12. Coe, R., Waring, M., Hedges, L.V., Arthur, J., *Research Methods and Methodologies in Education*. 2017, Los Angeles, CA: SAGE.
 13. Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*. 1988, Abingdon-on-Thames, UK: Routledge.
 14. Cohen J., *Things I have learned (so far)*. *Am Psychol*, 1990(45): p. 1304-1312.
 15. Cohen, L., Manion, L., Morrison, K., *Research Methods in Education*. 2018, London: Routledge.
 16. Creswell, J.W., *Qualitative Inquiry and Research Design: Choosing among Five Approaches*. 2007, Thousand Oaks, CA: Sage.
 17. Cronhjort, M., Filipsson, L., Weurlander, M., *Improved engagement and learning in flipped-classroom calculus*. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 2018. **37**(3): p. 113–121. doi:10.1093/teamat/hrx007.
 18. Day A.L., *Case study research*, in *Research Methods & Methodologies in Education, 2nd ed.* R. Coe, M. Waring, L.V. Hedges, J. Arthur, Ed. 2017, Los Angeles, CA: SAGE. p. 114-121.
 19. Duncan, D., *Clickers in the Classroom: How to Enhance Science Teaching Using Classroom Response Systems*. 2005, San Francisco, CA: Pearson Education.
 20. Dunn, P.K., Richardson, A., McDonald, C., Oprescu, F., *Instructor perceptions of using a mobile-phone-based free classroom response system in first-year statistics undergraduate courses*. *International Journal of Mathematical Education in Science and Technology*, 2012. **43**(8): p. 1041-1056. doi:10.1080/0020739x.2012.678896.

21. Dunn, P.K., Richardson, A., Oprescu, F., McDonald, C., *Mobile-phone-based classroom response systems: Students' perceptions of engagement and learning in a large undergraduate course*. International Journal of Mathematical Education in Science and Technology, 2013. **44**(8), p. 1160-1174. doi:10.1080/0020739x.2012.756548
22. Google, *Create forms*. 2017. Retrieved from <https://www.google.com.au/forms/about/>
23. Higher Education Research & Development Society of Australasia, *HERDSA Fellowship Scheme Handbook*. 2014, Milperra, NSW: HERSDSA. Retrieved from https://www.herdsa.org.au/sites/default/files/Fellowship\%20Handbook_6_5_2014.pdf
24. Johnston, B.M., *Implementing a flipped classroom approach in a university numerical methods mathematics course*. International Journal of Mathematical Education in Science and Technology, 2017. **48**(4): p. 485-498. doi:10.1080/0020739x.2016.1259516.
25. Jungic, V., Kaur, H., Mulholland, J., Xin, C., *On flipping the classroom in large first year calculus courses*. International Journal of Mathematical Education in Science and Technology, 2015. **46**(4): p. 508-520. doi:10.1080/0020739x.2014.990529.
26. King, S.O., Robinson, C.L., *'Pretty lights' and maths! Increasing student engagement and enhancing learning through the use of electronic voting systems*. Computers & Education, 2009. **53**(1): p. 189-199. doi:10.1016/j.compedu.2009.01.012.
27. Kline, R.B., *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. 2004, Washington, DC: American Psychological Association, p. 95.
28. Larkin, K., Calder, N., *Mathematics education and mobile technologies*. Math Ed Res J, 2016. **28**(1): p. 1-7. doi:10.1007/s13394-015-0167-6.
29. Lomen, D.O., Robinson, M.K., *Using ConcepTests in single and multivariable calculus*, in *Electronic Proceedings of the Sixteenth Annual International Conference on Technology in Collegiate Mathematics*. 2005. Retrieved October 17, 2017 from <http://archives.math.utk.edu/ICTCM/i/16/S107.html>
30. Love, B., Hodge, A., Grandgenett, N., Swift, A.W., *Student learning and perceptions in a flipped linear algebra course*. International Journal of Mathematical Education in Science and Technology, 2014. **45**(3): p. 317-324. doi:10.1080/0020739x.2013.822582.
31. Maciejewski, W., *Flipping the calculus classroom: an evaluative study*. Teaching Mathematics and its Applications, 2016. **35**(4): p. 187-201. doi:10.1093/teamat/hrv019.
32. Markie, P., *Rationalism vs. Empiricism*. 2017. Retrieved from <https://plato.stanford.edu/entries/rationalism-empiricism/>
33. Murphy, J., Chang, J., & Suaray, K., *Student performance and attitudes in a collaborative and flipped linear algebra course*. International Journal of Mathematical Education in Science and Technology, 2015. **47**(5): p. 653-673. doi:10.1080/0020739x.2015.1102979.
34. Naccarato, E., Karakok, G., *Expectations and implementations of the flipped classroom model in undergraduate mathematics courses*. International Journal of Mathematical Education in Science and Technology, 2015. **46**(7): p. 968-978. doi:10.1080/0020739x.2015.1071440
35. von Neumann, J., *The mathematician*, in *Works of the Mind*, R.B., Haywood Ed. 1947, Chicago: University of Chicago Press. p. 180-196.
36. Novak, J., Kensington-Miller, B., Evans, T., *Flip or flop? Students' perspectives of a flipped lecture in mathematics*. International Journal of Mathematical Education in Science and Technology, 2017. **48**(5): p. 647-658. doi:10.1080/0020739x.2016.1267810.

37. Center for Educational Research and Innovation, *Giving Knowledge for Free: The Emergence of Open Educational Resources*. Retrieved from <http://www.oecd.org/edu/cei/38654317.pdf>
38. Petrillo, J., *On flipping first-semester calculus: a case study*. International Journal of Mathematical Education in Science and Technology, 2016. **47**(4): p. 573-582. doi:10.1080/0020739x.2015.1106014.
39. Robson, L., *Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries: How to Show Whether a Safety Intervention Really Works*. 2001, Cincinnati, OH: DHHS.
40. Salzer, R., *Smartphones as audience response systems for lectures and seminars*. Analytical and Bioanalytical Chemistry, 2018. **410**: p. 1609-1613. doi:10.1007/s00216-017-0794-8.
41. Sawilowsky, S., *New effect size rules of thumb*. Journal of Modern Applied Statistical Methods, 2009. **8**(2): p. 467-474. doi:10.22237/jmasm/1257035100
42. Shadish, W.R., Cook, T.D., Campbell, D.T., *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. 2001, Belmont, CA: Wadsworth Cengage Learning.
43. Sullivan, G. M., Artino Jr., A.R., *Analyzing and interpreting data from Likert-type scales*. Journal of Graduate Medical Education, 2013. **5**(4): p. 541-542.
44. Tisdell, C.C., *Critical perspectives of pedagogical approaches to reversing the order of integration in double integrals*. International Journal of Mathematical Education in Science and Technology, 2017. **48**(8): p. 1285-1292. doi: 10.1080/0020739X.2017.1329559.
45. Tisdell, C.C., *Pedagogical alternatives for triple integrals: moving towards more inclusive and personalized learning*. International Journal of Mathematical Education in Science and Technology, 2018. **49**(5): p. 792-801. doi: 10.1080/0020739X.2017.1408150.
46. Tisdell, C.C., *On Picard's iteration method to solve differential equations and a pedagogical space for otherness*. International Journal of Mathematical Education in Science and Technology, 2019. **50**(7): p. 788-799. doi: 10.1080/0020739X.2018.1507051.
47. Tisdell, C.C., *Schoenfeld's problem-solving models viewed through the lens of exemplification*. For the Learning of Mathematics, 2019. **39**(1): p. 24-26.
48. Trochim, W.M.K., *The Research Methods Knowledge Base*. 2006. Retrieved from <https://www.socialresearchmethods.net/kb/positvsm.php>
49. University of Edinburgh. *What is digital education?* 2019. Retrieved from <https://www.ed.ac.uk/institute-academic-development/learning-teaching/staff/digital-ed/what-is-digital-education>
50. Wang, V.C., *Handbook of Research on e-learning Applications for Career and Technical Education: Technologies for Vocational Training*. 2009, Hershey, PA: IGI Global.
51. Wasserman, N., Norris, S., Carr, T., *Comparing a "flipped" instructional model in an undergraduate Calculus III course*, in *Proceedings of the 16th Annual Conference on Research in Undergraduate Mathematics Education*; S. Brown, G. Karakok, K.H. Roh, M. Oehrtman Ed. .2013, Denver, CO.
52. Yin, R.K., *Case Study Research: Design and Methods*. 4th ed. 2009, Thousand Oaks, CA: Sage.