Jl. of Computers in Mathematics and Science Teaching (2018) **37**(4), 1-32

A Pedagogically Effective Use of an Audience Response System to Increase Outcomes in Mathematics

ELIZABETH MCALPIN, DAVID SHILANE,
AND SELIN KALAYCIOGLU
New York University, USA
em341@nyu.edu
ds5548@nyu.edu
sk168@nyu.edu

Large foundational courses delivered in lecture halls are disadvantaged in eliciting participation from all students and maintaining engagement throughout the class session. We introduced a question-driven method facilitated by an audience response system, namely TopHat™, that allows the instructor to engage students in solving Calculus problems across the entire class as well as provide uniform immediate feedback to both the students, and the instructor as an opportunity to reflect and remediate. In Fall 2016, we compared outcomes for students who solved problems facilitated by TopHat™ in one section (041) to those who solved problems using paper in another section (046). Results from statistical tests found the section (041) using a question-driven method facilitated by TopHat™ was associated with statistically significantly improved performance measures resulting in a 4.98% increase in final grades amounting to a half a letter grade difference, compared to the control section (046) solving problems facilitated with paper. The improvement in performance was driven by statistically significant differences on the homework assignments and midterm examinations in the section (041) using TopHat™. The TopHat™ section (041) also scored higher on the final exam although not statistically significant average compared to section (046). Greater performances scores contributed to statistically significant final grade averages.

Keywords: *question-driven method, audience response system, active learning, interactive instruction, technology-mediated instruction, higher education, mathematics, calculus*

BACKGROUND

Calculus I, the largest enrolled undergraduate course at New York University, provides about 14 sections each year with over 100 students in each section. Close to 1,400 students fill all 14 sections each year. In the past, the course delivered in-class lecture demonstrations followed by homework to be completed outside of class. As of Fall 2015, in an effort to introduce more time for student-centered activities and decrease instructor-centered demonstration 2 sections out of 14 experimented with a "flipped" model. The flipped teaching method presents an alternative design to the traditional attend lecture then completed homework sequence. In the flipped approach, part of the in-class lecture content and homework is accessed before class, thus, freeing in-class time to provide more student-centered problem solving activities. In this model, one class period is dedicated to lecture and one class period a week is dedicated to students work in small groups (or individually) to solve additional homework problems. The instructor (and teaching assistants) provides real-time guidance to students as requested. Even though the flipped model statistically improved overall grades compared to the traditional format we wanted to understand more about the specifics of the model.

In the Spring of 2016 we observed the two Calculus I sections using the flipped model to better understand the student experience in this new format. Students in both sections attended two 75-minute class periods each week. The first scheduled 75-minute class delivered instructor-led demonstrations solving Calculus problems. The faculty member used ample questioning during the demonstrations to engage the students in the process. We observed the more active students sat in the front and raised their hands to answer questions more often than students sitting in the back of the lecture hall. Students in the back of the class were less engaged as observed by texting on wireless devices, sleeping, or focusing on other content not related to the course. The second 75-minute class focused exclusively on student-centered problem solving time. Teaching assistants and instructor navigated the space fielding questions from students as they worked in small groups or individually to solve homework problems due at the end of week. Interactions between the student and expert were personalized based on student needs. Students reported enjoying the second class more than the first class due to the time dedicated to solve problems with as needed personalized help from the experts. Even though most students were engaged in solving problems either in small groups or individually we still observed some off-task behaviors such as texting, internet surfing, or personal conversations. We also counted a progressive drop in attendance over the semester. Despite the ample assistance from teaching assistants and instructor, managing the classroom space with all the students proved challenging to meet the needs of students. We questioned if there was a more effective and efficient way to provide student-centered problem solving in both class periods to engage a larger audience more uniformly. Likewise, we questioned if the engagement solution might also improve outcomes.

Thus, the goal of this research was to investigate whether a question-based method facilitated by an audience response system such as TopHat™ improves student engagement and performance outcomes compared to a massed problem solving method facilitated by paper to solve the same in-class homework problems. It should be noted that in either case the instructor highly encouraged students to work collaboratively to solve problems although this recommendation was not strictly enforced. Students could also choose to work individually.

To this end, in Fall 2016, we compared the learning outcomes of students in section (041) solving problems using TopHat™ to the students in section (046) engaged in problem solving without TopHat™. After students enrolled into the two sections targeted for this study, we randomly assigned the two sections to treatment groups; TopHat™ or no TopHat™. To investigate whether the use of this technology improved outcomes we analyzed various behaviors and performance outcomes such as weekly quizzes, midterm scores, final exam scores, and final grades.

RESEARCH ON AUDIENCE RESPONSE SYSTEMS TO SUPPORT ACTIVE LEARNING STRATEGIES

A. Introduction to Audience Response Systems

To overcome participation, engagement, and attendance issues in large format courses, audience response systems (ARS) present *one way* to improve these issues. An ARS can be used to facilitate formative assessment and active learning pedagogy in the classroom or online to build valued skills and knowledge. It allows students to respond to questions presented in various formats (e.g., multiple choice, fill-in response), displayed through the instructor's presentation slides, and projected on a screen in a classroom. Students register their responses using a wireless devise (e.g., phone, tablet, laptop) through a web browser. The results are automatically collected in real-time, summarized and presented to the class in some graphical format (e.g., bar chart). Responses are anonymous to peers yet accessible by the instructor through an online administrative portal allowing the instructor to measure and analyze student responses over time. They are commonly used in large format courses and can be used individually or in small groups (Daniel & Tivener, 2016). This kind of click to respond technology may be referenced under various terms such as Clicker, Electronic Voting System (EVS), Audience Response System (ARS), Personal Response System (PRS), and Student Response System (SRS), to name a few (Cain & Robinson, 2008; Kay & LeSage, 2009). For the purpose of this paper we will use the term audience response system, ARS, for further discussion.

B. Theoretical Rationale for A Question-Driven Method with an ARS

The main perceived benefit and purpose of ARS is to promote cognitive and non-cognitive goals such as active learning, formative assessment, and contingent teaching in the classroom. For example, to overcome low attention issues found in students during the delivery of long lectures, Kay and LeSage (2009) cite research literature that supports breaking up long lecture classes into 20 minute intervals to present questions facilitated by an ARS. Studies comparing learning performance between this kind of active learning technique to traditional methods that deliver long lectures with little interaction between instructor and student demonstrated success (Bullock et al. 2002; El-Rady, 2006; Schackhow et al., 2004; Slain et al., 2004) in part because student attention diminishes after 20 minutes or less (d'Inverno et al. 2003; Jackson, Ganger, Bridge, & Ginsberg, 2005) and there is no opportunity for students to articulate thinking nor contingent teaching to remediate misconception. Students learn more when they participate in their process of learning, whether it's through discussion, practice, review, or application, to name a few strategies (Grunert, 1997; Prince 2004; Mayer & Wittrock, 2006; Kay and LeSage, 2009). Lectures and demonstrations are important for conveying domain-specific information. Enhancing lectures and demonstrations with interactive opportunities allows students to immediately apply the instructional message and receive timely feedback. It is important to note that when choosing an ARS to support cognitive and non-cognitive goals, an ARS promotes a question-driven model.

A question driven model supports student learning through generative learning (Mayer & Wittrock, 2006; Wittrock, 1990). In generative learning students engage in active cognitive processing when primed by an instructional method. There are many instructional methods from which to choose. Posing questions or problems to solve is one method. Questions may also vary in terms of type such as multiple choice or open response. Questions may be interleaved between lecture demonstrations or massed at the end of a lecture. They may also vary from retention based to deeper understanding to support learning objectives. The method may be successful depending on how the questions are constructed. Duncan (2005) suggests questions should be aligned with valued learning objectives such as demonstration of mastery and reasoning. Questions should also present a challenge to students (Duncan, 2005). Questions should also have well-constructed distractor answers that target common misconceptions and errors so that remediation is specific to each distractor (Banks, 2003). Despite these feature differences around the method, it is the method that *may* cause learning not the media itself (Mayer, 2009). Employing a well-researched method or developing a new one helps instructors evaluate their own teaching and learning practices and protects against ad hoc practices (Schoenfeld, 2002). A method may be afforded by a technology (Mayer, 2009).

A major goal of a question-driven model is to promote more interactive communication between the instructor and the majority of students in the classroom but the question-driven model can take many forms (Mayer, 2009; Lantz & Stawiski, 2013; Stewart & Stewart, 2013; Morais et al., 2015). Success of the method may be dependent on the teaching routine, namely, the placement of questions in the instruction. The teaching routine could present a question before instruction, interleave questions with instruction or mass questions at the end of instruction (Mayer, 2009; Lantz & Stawiski, 2013; Stewart & Stewart, 2013; Morais et al., 2015). Stewart & Stewart (2013) discuss an interactive contingent teaching model using an ARS that begins with an introduction to the topic followed by some kind of student-centered activity (i.e., group discussion, case study, or experiment) followed by planned instructional pathways depending on how students respond. Once the model reaches its final conclusion the sequence repeats to address the next instructional topic. In this way, the questions are interleaved with instruction and may even drive the instructional path. While the study conducted by Stewart & Stewart (2013) examined the students' and instructor's perceived use of the new model enhanced with an ARS with positive feedback, the researchers advocated more in depth research investigating the impact on student learning.

From a pedagogical perspective, a question-driven approach may be supported by generative theory (Mayer 2001, Renkl & Atkinson, 2007, Wittrock, 1990) whereby the learner is at the center, actively engaged in the assimilation, integration, construction, and application of knowledge as opposed to considering learning as merely the transmission of knowledge via a lecture, textbook, or video.

A question-driven model, with or without technology, requires breaking long instruction time into smaller parts to add valued questions in between parts. When the delivery of instruction is broken into smaller parts that focus on teaching one concept or skill at a time, students can more easily process and apply the instructional message (d'Inverno et al. 2003; Jackson, Ganger, Bridge, & Ginsberg, 2005). If the delivery of the instructional message contains too many steps or pieces of information for students to retain in working memory, students will drop essential information or experience cognitive overload impairing learning (Sweller, 1988). Allowing students to practice immediately after receiving the short instructional message to be learned allows students to reflect on the message and integrate it into long-term memory (Skuballa et al., 2018).

To support a question driven model, with or without technology, feedback and adjustment are part of the model. Chien et al. (2016) propose that a possible strength of using an ARS is to promote ample "performance-feedback-adjustment-loops" for every student engaged in the question-driven model. Shute (2008) defines formative feedback "as information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning" (p. 152). Formative feedback has different implementation features including; correct/incorrect results, immediate versus delayed timing, explanation versus no explanation on results. Immediate feedback allows students to evaluate themselves and adjust learning gaps between their current state and the learning goal (Shute, 2008). Azevedo & Aleven (2013) suggest that performance-feedback-adjustment loops that provide explanations beyond correct/incorrect results can facilitate metacognitive awareness which helps students to become more effective in their learning process. According to Chien et al. (2016), the "performance-feedback-adjustment loop is difficult to iteratively implement in the classroom without an efficient and reliable tool for data collection and calculation" (p. 5). This statement implies that an ARS can reduced administrative and classroom management inefficiencies that may have otherwise reduced the effectiveness of learning.

Lastly, the privacy enabled by an ARS might be helpful to increase participation especially from lower performers who may feel inhibited to answer questions publically for fear of failure or peer judgment (Chient et al., 2016). Threats to the *self* can reduce motivation and participation resulting in poor performance (Hattie & Timperley, 2007). As observed in the Calculus I classroom, students who sat in the front of the class were more likely to raise their hand to answer questions than students in the middle or back of the room. We perceived students sitting in the front to raise their hand more because they had more self confidence that their responses would be correct and receive positive reinforcement from the instructor. Thus, the privacy feature of an ARS may reduce student inhibitions and encourage more participation from students sitting further away from the front of the classroom.

To summarize the theoretical rationale, a question-driven method facilitated by an ARS may impact learning positively by improving classroom management, engaging more students to participate due to anonymity, and provides immediate feedback allowing for more self-regulation and instructor remediation opportunities. The success of the question-driven method facilitated by an ARS may also depend on the placement of questions during instruction. Interleaving questions (along with feedback and explanations) with the core lecture demonstrations may be more impactful than massing questions (along with feedback and explanations) at the end of a long lecture demonstration.

C. ARS in Varied Disciplines

Literature supporting the use of ARS report improved student engagement in-class (Draper and Brown, 2004), increased or sustained attendance (Bullock et al., 2002; Mastoridis & Kladidis, 2010), provided valuable feedback to students and instructor allowing for just-in-time instructional remediation (Caldwell, 2007), and increased learning performance (El-Rady, 2006; Mayer et al., 2009) and knowledge retention (Terenzini et al., 2001; Dhaliwal et al., 2015). Using an ARS does not preclude students from choosing to work in pairs or individually to solve problems. Rather, it can enhance the learning whether working with peers or individually (Daniel & Tivner, 2016).

However, in a meta-analysis by Hunsu et al. (2016) report only a few robust empirical studies comparing ARS facilitated learning environments to non-ARS facilitated learning environments on different cognitive measures such as course grades, retention, and transfer. The authors report that most studies comparing students in ARS and non-ARS classrooms within a variety of disciplines did not establish a baseline to determine similarities or differences between groups prior to the intervention. Factors like these undermine the validity of purported results. The results of this meta-analysis suggest no significant differences between mean scores comparing ARS to non-ARS groups on retention (recall) types of questions. This finding supports the claim found by Mayer & Chandler (2001) that substantive effects are not found in the use of multi-media learning when using measurements focused on retention but can be found when using measurements focused on transfer or deep-understanding. Deep understanding may be facilitated through reflective strategies such as think-pair-share combined with technology such as ARS (Hunsu et al., 2016).

Mayer et al. (2009), using a robust experimental design, examined the effectiveness of an ARS to implement a questioning method in a large lecture format class in educational psychology. The study compared an experimental condition using a question-driven model facilitated by an ARS to a condition using a question-driven model without an ARS as well as a control group who experienced no in-class questioning or ARS. Controlling for prior ability the ARS questioning group improved academic achievement by 1/3 grade point over the 2 other conditions. Mayer (2009) found that the no ARS group had no effect on learning outcomes compared to the control condition, no ARS and no questioning. Mayer (2009) suggested that the questioning method worked with the ARS technology because it facilitated higher cognitive engagement, deeper cognitive processing and reduced administrative and classroom management inefficiencies that may have otherwise reduced the effectiveness of learning.

While the study aimed to compare two groups using the same question driven method in order to isolate the use of technology there were subtle differences in terms interleaving questions with lecture or massing questions at the end of lecture. Mayer et al. (2009) reported that the ARS group received questions at the end of a section of the lecture while the non-ARS group received the same questions but distributed “at the end of the lecture (or section of lecture) (p. 54).” This small difference, questions interleaved by section of lecture versus massed at the end with lecture, may have impacted students cognitive processing abilities of the lecture material differently. Mayer, as instructor and researcher of this study, alluded to the idea that the question-driven method could not be implemented successfully without the technology. He conjectured the question-driven method facilitated by paper may not have worked due to logistics that caused more disruption in the class than the class using technology. The reason for the difference may be due to the affordances of the technology to more efficiently facilitate the question-driven method that interleaves questions so that students could more easily process the material and practice one concept at a time. Interleaving with an ARS may have facilitated a more impactful performance-feedback-adjustment loop. Mayer et al. (2009) admitted to massing the questions as the end of the lecture "which involved 'quiet time' and may have created too much separation from the lecture" (p.55). He also observed less participation to answer follow-up responses. More disruption, longer separation time from instructional content, and less participation in the question-driven method facilitated by paper may have detracted from the effectiveness of the method.

Lantz & Stawiski (2015) found no effect between distributing questions via an ARS in between instruction covering topics in an undergraduate psychology course compared to massing questions at the end of instruction with an ARS. The authors admit that the instructional video used was much shorter in overall duration and segmented into three distinct topics compared to the typical length of a lecture that covers many more topics in class. Further research is needed to compare these two question-driven methods.

The use of an ARS comparing other interactive or question-driven instructional methods on outcomes favors conditions using an ARS along with a question-driven model compared to groups using a different model along with an ARS or without an ARS (Hunsu et al., 2016). Studies comparing the effective use of an ARS to non-ARS using similar question-driven strategies yield negligible differences. According to Hunsu et al. (2016), the effect of using an ARS disappears when both ARS and non-ARS conditions used the same question method. Authors of this paper question the specifics to this claim especially when combining other dimensions into consideration such as class size, topic domain, question placement (e.g., interleaved or massed), and other details surrounding the question method.

Aside from the studies comparing strategies used with clickers, according to Hunsu et al. (2016), the classroom size matters. The authors of this meta-analysis suggest a small benefit to using an ARS in larger class sizes. Again, authors of this paper question whether only looking at class size is a fair evaluation when other factors may contribute to the results such as specific question method, topic domain, and question type.

D. ARS in Mathematics in Higher Education

Many studies covering subjects related to Mathematics in higher education focused on examining attitudes and perceptions towards the use of an ARS rather than learning outcomes (Dunn et al., 2013; Stewart & Stewart, 2013; Morais et al. 2015). Since this paper focuses on learning outcomes the discussion of the literature turns towards studies contributing results on learning outcomes. Literature supporting the use of ARS to improve learning outcomes specifically in the field of mathematics for higher education yielded mixed results.

In an effort to improve student retention, participation, and learning outcomes, Liu and Stengel (2009), investigated the use of an ARS in college level courses covering topics in statistics. The instructor routine usually presented a sequence that commenced with an introduction of some basic concepts and worked example demonstrations followed by a "multiple choice" question to solve a problem supporting the instruction. Students submitted responses via clickers. A distribution chart displayed to all students their responses whereby the instructor could decide to remediate with another similar question or move onto the next topic. The researchers compared two sections using this question-driven method with an ARS to two sections using formal lecture along with some in-class activities not using an ARS. Comparable sections used the same in-class questions, homework problems, weekly quizzes, and exams. Liu and Stengel (2009) found statistical significance for the sections using the question-based method with an ARS for student retention and learning outcomes. The authors suggest that an ARS works best for complex problems involving several steps because the questions can be delivered in steps rather than one large problem. Students can then be assessed and remediated at each step level allowing for improved understanding of the connections between the steps as well as making small, just-in-time, adjustments to their learning. While the study provided positive results, it did not control for prior knowledge measures. There is also some confusion around the authors only including test results for students who completed the course in order to reduce a bias in favor of sections not using the clickers. Future studies should aim to improve some of these validity issues.

Morais et al. (2015) investigated the use of Problem Based Learning (PBL) facilitated by and ARS to teach a large undergraduate course in Mathematics. The researchers compared an experimental group using a PBL model facilitated by an ARS to a control group who received the conventional teaching method used over prior years. The sections were facilitated by different instructors. The study results yielded significant differences between the conditions favoring the PBL model facilitated by an ARS over the conventional teaching model. While the study provided details about the PBL-ARS model it did not provide any details about the conventional teaching model. It is unclear if the PBL-ARS model provided more practice opportunities related to the assessment compared to the conventional teaching model. There is no indication that the same questions used in the PBL model were used in both conditions to compare strategies or to control for this factor. The comparison is unclear.

In a study by Butler et al. (2016), researchers compared three conditions for a large enrollment undergraduate course in Algebra; the use of power points slides and ARS to facilitate a question driven model; power point slides with no ARS to drive questions; and a traditional lecture with no power point slide or ARS. The study results yielded no significant differences between the three conditions. Since different instructors were assigned to a condition, the differences in teaching ability and style may have confounded these results.

Due to variations in study method designs and validity issues the literature on the effective pedagogical use of an ARS demonstrates mixed results within the discipline of mathematics in higher education. Thus, the present study aims to fill gaps from prior studies to provide a more ecologically and methodologically valid design by using an authentic classroom setting, a baseline, the same questions in both sections, the same instructor in both sections, and use over the entire semester as opposed to one shot to reduce novelty effect. We adopted the coding scheme from Chein et al. (2016) to explain feature similarities and differences of the study (*Appendix: Table 1)*. We added a few extra features such as instructor, time of day, immediate feedback, and points.

There are two main differences between the two sections. One difference is the question method and the other is the use or non-use of an ARS to support the question method. The ARS section used a question-driven method whereby the questions were interleaved with instruction and remediated explanations during both weekly scheduled classes. The questions aimed to test students on how well they understood the demonstration or instructional message. Questions were crafted with distractors that indicated misconceptions. Students result aimed to inform the instructor which misconceptions to remediate with further demonstrations before moving onto the next concept.

The non-ARS section used a massed question method during the second 75-minute class. The decision for the massed question method was mainly to address classroom management issues using paper to answer questions in a large enrolled section. It was important for the instructor to complete required instruction time without delays or confusion otherwise caused by a paper based administration of questions. Attempting to interleave paper-based questions with demonstrations would have delayed instruction time. It was also important to aim to engage as many students as possible in the activity. If the instructor decided to forgo collecting student results to save time participation incentives may drop because students would not be held accountable. To call on students for the correct answer would have yielded the same participation issue whereby students in the front of the class are more likely to answer. One challenge with an ARS is deciding how much time to give students to work on problems. There will be variation in the speed with which students need to complete questions. More advanced students will complete questions faster than less adept students. In the massed method allows for more individualized instruction where students can work at their own pace and seek personalized help as needed. There will always be benefits and risks to implementing a method, the question is which one supersedes the other.

E. A Question-Driven Model for ARS based Instructional Methods: Rationale for the Present Study

The present study supports the rational for a question-driven model that interleaves questions with demonstrations facilitated by an ARS for a large format introductory course in mathematics because: (a) the instructional material is delivered in smaller parts with conceptual and problem solving questions delivered after each parts allowing students to more easily process the instructional message and immediately apply it before moving onto the next concept, (b) facilitates a performance-feedback-adjustment loop for all students promoting self-regulation and metacognition, (c) greater participation due to the privacy feature (d) students may be more engaged to listen attentively during instruction because they know a question will soon appear, (e) all students receive the same explanations for incorrect responses allowing for more uniform instruction across the class, and (f) reduces classroom management issues that may impair learning.

The present study aims to evaluate behavioral and performance measures to determine if using a question-driven method and an ARS, TopHat™, during class to solve weekly homework problems (individually or with peers) improves student engagement and performance outcomes compared to a massed-question method using paper.

We propose the following hypothesis; the section (041) using TopHat™ in class to tackle homework problems will perform better on performance measures including overall course grades than the section (046) using the massed question method in class without the use of the ARS technology to complete the same homework problems. Our hypothesis is supported by the theoretical rationale and features outlined above.

METHODS

Study design

In order to evaluate courses for improvement, faculty may test the effectiveness of a new way of teaching in a section before introducing the new method to other sections. A quasi-experimental design was used for this project and conducted by an internal research team within the university. At New York University, undergraduate students participated in this project during regular scheduled instructional class time for the Calculus I course. The same instructor facilitated both sections and contributed to this study as final editor. Students self-enrolled into either section of the course using normal enrollment processes. The sections were randomly assigned to either the use of TopHat™ or no TopHat™. To establish a baseline and control for any prior knowledge differences, we included the following measures SAT/ACT scores, Grade Point Average (GPA), and Diagnostic. The Institutional Review Board (IRB) approved this project.

Participants

Undergraduate students enrolled in Calculus I must meet one prerequisite (e.g., SAT score of 650 or higher). *Table 2* describes the distribution of the sample size based on the total number of students in both sections. A combined 231 students enrolled in the course. Section (041) (TopHat™) had a total of 128 students while Section (046), non-TopHat™, had a total of 103. Section (041) had a female to male ratio of 1. Section (046) had a female to male ratio of 1.36. Overall both groups had a female to male ratio of 1.15. Majors and non-majors could not be determined since many students had not yet declared majors. Students with declared majors come from a wide range of over 50 undergraduate disciplines.

Table 2
Participants

	Section (041), ARS -TopHat™	Section (046), Non- TopHat™
Total N = 231	128	103
Males	61	42
Females	61	57
N/A	6	4

Study Procedure, Materials, and Apparatus

A quasi-experimental design was used for this project where we compared outcomes on weekly homework, exams and final course grade.

Both sections of the course completed the same diagnostic, weekly online pre-work modules, weekly homework, weekly quizzes, mid-term and final exams. Both sections attended two 75-minute face-to-face classes throughout the semester. Both sections met on the same days during the week but at different times. Section (041) met in the afternoon while section (46) met in the morning. Additionally, students attended one 75-minute recitation period managed by teaching assistants who fielded questions from students as well as administered weekly paper quizzes.

The homework questions used in-class for both groups were much more challenging than questions on end-of-week quizzes and more like exam questions. In many cases the homework questions were more difficult than exam questions and targeted deeper understanding (*Appendix: Image 1: Sample Homework Problem on Paper*). Both conditions had to complete the paper-based version of the homework for final grading and feedback at the end of the week. The majority of questions on the paper version required open, fill-in, responses. Final homework submissions required not only the solution but the worked-out steps to the solution. Homework assignments were worth 10% of the final grade.

Both sections had at least 4-6 different teaching assistants assigned to grade end of week homework, facilitate recitation, administer weekly quizzes. Both sections used the same online lessons, paper-based homework, weekly paper-based quizzes, and paper-based exams.

The materials for section (041) using TopHat™ consisted of PowerPoint slides projected on a large whiteboard for both 75-minute weekly class periods for the entire semester. The instructor used an electronic writing pad to write out her demonstration on top of the slides dedicated to lecture demonstrations. The lecture demonstration slides were interleaved with student-centered problem solving questions. This allowed the instructor to integrate the lecture demonstrations with the student-centered questions with minimal disruption to the class flow. During demonstrations of problems she used a think aloud and Socratic questioning method to work through the material. Often the instructor would engage students to openly answer questions as she worked through the demonstration._Lecture demonstration slides were broken up by student-centered practice problem solving questions in the form of multiple choice or fill-in. Open, fill-in, response was only used when the value could be assessed without typing error. Students used their phone, laptop, or tablet to submit a response to the question on display. Students could work with another student or individually to solve the problems but were encouraged to work with a peer. Students were given about five minutes to complete a given problem. Once the poll closed the responses were automatically aggregated and displayed on the screen as a bar chart with percentages. The instructor remediated with in-depth explanations to incorrect answer submissions before moving on to the next concept. The ARS software collected student university email addresses in order to collect attendance and participation data. Students were not penalized for answering a question incorrectly. The homework questions along with immediate feedback and remediation provided enough incentive to participate in-class with the ARS.

For the TopHat™, section (041), some of the more challenging homework questions containing several steps were delivered in a sequence of smaller problems building up to the whole problem (*Appendix: Image 2: Sample Homework Problem Displayed in Parts Using TopHat™).* Delivering the challenging homework questions in a set of smaller questions acted as form guided scaffolding. Students received immediate feedback and explanation at each step of the larger problem. Students transferred the homework problems completed in-class using the ARS to the paper version required for final grading and feedback at the end of the week.

The non-TopHat™ section (046) only used an attendance tracking software connected to Google™ spreadsheets during class to collect attendance at the beginning and end of class. The materials for the non-TopHat™ section consisted of the same PowerPoint slides projected on a large whiteboard for one 75-minute weekly class periods for the entire semester. The 75-minute class time was presented and delivered in the same interactive, Socratic way as the TopHat™ section except for the embedded questions. The second 75-minute class was devoted to student problem solving time using the same homework problems except on paper. Students used the paper version required for final grading at the end of the week. Students were encouraged to work in pairs or small groups but could also work individually. The instructor and teaching assistants attended to student questions as they worked to solve the homework problems on paper during the class period. Generally, the homework questions along with as needed guidance provided enough incentive to participate in the in-class activity.

Data Analysis

This study makes use of a quantitative analysis including summary of behaviors and performance, correlations, and a regression model on course grades by section and prior measures.

Measures:

We analyzed various measures of student behaviors and performance in the course. Following is a list and definition of the variables grouped as indicators of prior, behavior, and performance measures:

Prior Knowledge

1. *SAT/ACT Math score*: math score on either Scholastic Aptitude Test (SAT) out of 800 points or American College Test (ACT) standardized tests out of 36 points (converted to 800-point scale)
2. *Current GPA*: the current Grade Point Average (out of 4 points)
3. *Pretest Diagnostic*: Score out of 42 from an algebra test administered at the beginning of the course was used a proxy for students' prior math ability (converted to a 0 to 100 percentage)
4. *Section*: the section in which the students enrolled by choice (with or without TopHat™)

Behavior

5. *Attendance rate*: the average attendance rate of lectures attended by the student over the duration of the course (out of 26)
6. *Module rate*: the average completion rate of online lesson modules completed prior to attending class (out of 48)
7. *Homework rate*: the average completion rate of homework completed over the duration of the course (out of 11)
8. *Quiz rate.* the average completion rate of quizzes completed over the duration of the course (out of 12)

Performance

9. *Homework scores.*: average score across all the homework (11, each with varying total items, 10 points) in the course
10. *Quiz scores*: average score across all the quizzes (12, each for 10 points) in the course
11. *Midterm 1 scores*: score on the first midterm exam (out of 100 points)
12. *Midterm 2 scores*: score on the second midterm exam (out of 100 points)
13. *Final scores*: score on the final exam (out of 100 points)

Final outcome

14. *Course grade*: a weighted sum of attendance and lesson modules (5%), WebAssign™ problem sets [1] (5%), quizzes (10%), homework (10%), midterms (40%), and final (30%).

Study Hypothesis Question

What are the effects of an interactive, question-driven, teaching method facilitated by an audience response system (TopHat™) versus the current teaching method of massed lecture time and massed problem-solving time without an audience response system? In other words, does the section (041) with TopHat™ versus section (046) without TopHat™ improve performance measures and overall Course Grade?

Study Outcomes

The primary outcome measures were the increase in student performance measures namely average course grade but also included average weekly homework scores, average weekly quiz scores, midterm 1, midterm 2, and final exam.

Statistical Analysis

We investigated the course's data to understand the differences between the two sections. The variables measuring the students' prior knowledge, behaviors, performances, and final outcomes were explored with descriptive analyses. Univariate comparisons of each measure were performed using *t*-tests across the two sections. *Correlations* were drawn to determine the strength of the relationship between the variables and to investigate their suitability for use in regression models.

Linear Regression was used to model the final outcome in each section while controlling for the measures of prior knowledge. The difference in outcomes for each section was evaluated based on the statistical effect.

Results

Prior Knowledge Abilities

Descriptive statistics included the mean and standard deviations (SD) to show similarities and differences between the two groups in terms of prior knowledge abilities (*Table 3A*). Section (041) using TopHat™ demonstrated slightly higher gains in prior knowledge abilities but these differences were not statistically significantly different.

Descriptive statistics show similarities and differences on prior abilities between gender in each condition (*Appendix: Table 3B*). There was no a statistically significant difference on any prior ability (SAT/ACT Math, GPA, Pretest Diagnostic) between males and females in each condition. Since there were no differences between males and females on prior knowledge, we proceed with a unified analysis.

Table 3A

Summary of Prior Knowledge Abilities (All students by Section)

		TopHat™ Section (041) (*n= 128*)	NonTopHat™ Section (046) (*n= 103*)	*t-test Differences*
Indicators	*Measure*	*Mean (SD)*	*Mean (SD)*	*t-value, p-value*
Student Prior Abilities	SAT/ACT Math	729.6 (56.3)	719.7 (67.4)	1.146, (0.253)
	GPA	3.4 (0.5)	3.3 (0.5)	1.476, (0.141)
	Pretest Diagnostic	76.1 (14.9)	75.50 (15.0)	0.290, (0.772)

Behaviors

As shown in *Table 3B*, both sections applied, on average, about the same effort to attend class, complete weekly required work, and weekly quizzes aimed to prepare students for exams. Attendance rates were collected via TopHat™ for section (041) and by Google Attendance tracking software for section (046). As part of the course requirements students needed to complete online lesson modules and practice problems before attending the first scheduled class during the week. Completion rates on the online lesson modules were collected through the university's learning management system. Completion rates for end of week homework problems and quizzes were also collected and tallied. The mean rates suggest that both conditions were equally prepared with the course materials.

Table 3B

Summary of Behaviors (All students by Section)

		TopHat™ Section (041) (n= *128*)	NonTopHat™ Section (046) (n= *103*)	*t-test Differences*
Indicators	*Measure*	*Mean (SD)*	*Mean (SD)*	*t-value, p-value*
Student Behaviors	Attendance rate	89.0 (21.8)	87.0 (18.4)	0.738, (0.461)
	Module rate	93.0 (20.6)	95.3 (11.6)	-1.064, (0.288)
	Homework rate	95.0 (11.7)	94.2 (13.0)	0.499, (0.618)
	Quiz rate	99.5 (4.5)	99.7 (1.7)	-0.281, (0.779)

Initial Performance Differences

Statistics including the mean and standard deviations (SD) show similarities and differences between the two groups on per- formance measures (*Table 3C*). Univariate, two-sample *t*-tests were per- formed on each measure to compare the mean performance by section. We note, the TopHat™ section (041) demonstrated statistical significant differ- ence on Weekly Homework Scores, Midterm 1 and Midterm 2 compared to the NonTopHat™ section (046). These findings support the main prediction the question-driven method facilitated by an ARS would improve academic performance We note no statistically significant difference between sections for the weekly quizzes. Meanwhile, while section (041) using TopHat™ scored higher on the Final Exam compared to section (046) this difference was not statistically significant. The differences in performance measures cumulated to the overall course grade that also demonstrated a statistically significant difference in favor of the TopHat™ section (041). While there is an 8.1% difference in overall course grades we use linear regression to investigate the extent to which this effect would be attenuated when controlling for the measures of prior knowledge.

Table 3C
Initial Performance (All students by Section)

		TopHat™ Section (041) (*n=128*)	NonTopHat™ Section (046) (*n=103*)	*t-test Differences*
Indicators	*Measure*	*Mean (SD)*	*Mean (SD)*	*t-value, p-value*
Student Performance	**Homework scores**	**81.1 (11.2)**	74.7 (11.9)	4.169, (**0.00**)
	Quiz scores	81.4 (17.6)	77.9 (18.5)	1.484, (0.139)
	Midterm1 scores	**87.1 (11.6)**	79.4 (15.9)	4.125, (**0.00**)
	Midterm2 scores	**71.8 (18.3)**	66.3 (20.2)	2.135, (**0.034**)
	Final scores	79.2 (17.8)	75.8 (17.0)	1.443, (0.150)
Final Outcome	**Course grade**	**86.6(12.6)**	78.5 (14.8)	4.383**, (0.00)**

Correlations

The correlation results (*Table 4A*) between prior knowledge variables, performance measures and course grades show prior knowledge scores correlate significantly with performance scores and course grades despite the ranges from low to moderately high positive relationships. Diagnostic shows the lowest relationships among performance measures and course grades, especially, homework scores (Pearson $r = .27$, $p < .01$) and quiz scores (Pearson $r = .33$, $p < .01$). GPA shows the highest moderate relationship among performance measures and course grades (Pearson $r = .65$, $p < .01$). Homework average are most highly correlated with course grades (Pearson $r = .87$, $p < .01$) and quizzes (Pearson $r = .70$, $p < .01$) Although measures of prior knowledge have low to moderate correlations with all of the outcomes, correlations were relatively weaker for the Homework and Quiz average than for exams. This observation may suggest that an ARS such as TopHat™ can help to fill the gap in assignments and evaluations like quizzes. We also observe higher correlations between SAT/ACT Math Scores and testing performance on the Midterms and Final Exam than between SAT/ACT Math Scores and Homework Average or Quiz Average. This may suggest that prior knowledge indicators such as SAT/ACT Math Scores to be more impactful for predicting exam scores due to a combination of prior knowledge and test-taking abilities. GPA appears to be a moderately strong indicator for all performance variables including course grade. Students with higher scores in any of the prior knowledge variables tend to have higher performance scores and course grades.

Table 4A

Correlations between Prior Ability, Performance Measures, and Course Grades (overall sample, N=231)

	Diagnostic	*SAT/ACT Math Scores*	*GPA*	*Homework Average*	*Quiz Average*	*Midterm 1*	*Midterm 2*	*Final Exam*	*Course Grades*
Diagnostic	1								
SAT/ACT Math Scores	.28	1							
GPA	.25	.31	1						
Homework Scores	.27	.30	**.55**	1					
Quiz Scores	.33	.24	**.56**	**.70**	1				
Midterm 1	.40	**.49**	**.53**	.57	**.64**	1			
Midterm 2	.36	**.50**	**.59**	.51	**.57**	.73	1		
Final Exam	.35	**.48**	**.60**	.53	**.62**	.73	.75	1	
Course Grades	.37	**.45**	**.65**	**.87**	**.77**	.80	.75	.81	1

Note. Correlation is significant at the 0.01 level (2-tailed).

Linear Regression: Modeling Course Grades by Section and All Prior Measures (Diagnostic, GPA, SAT/ACT)

The linear regression model shows course grades by section and all prior measures (*Table 5*). The p-value of the t-test for section contributes to the model as well as each prior measure predictor. The difference between sections is significant while adjusting for the measures of prior knowledge (p= 0.000). For students in section (041) using TopHat™, predicts a 4.98 point score higher than students in section (046). It appears that the small effect gains from the homework and exams accumulated to a larger effect gain in overall course grade over and above prior knowledge predictors. Meanwhile, the prior measure scores independently also have a positive impact, with higher prior measure scores significantly associated with higher Course Grades (p = 0.000). This that prior knowledge and test taking abilities predict student success. The model explains about 53% of the variation in Course Grades, which was a statistically significant amount *F* (4, 196) = 55.549, p = 0.000. These findings support the main prediction the question-driven method facilitated by an ARS would improve academic performance.

Table 5

Modeling Course Grades by Section and All Prior Measures (N= 200)

Dependent Variable	*Source*	*Unstandardized Coefficients*		*Standardized Coefficients*	*t value*	*p value*
		B	*Std. Error*			
Course Grade	Intercept	13.498	8.721		1.548	(0.123)
	Section (041)	4.98	1.392	0.877	3.577	(0.00)
	Diagnostic	0.197	0.052	0.197	3.8	(0.00)
	GPA	14.585	1.496	0.514	9.746	(0.00)
	SAT/ACT	0.047	0.012	0.203	3.832	(0.00)

Linear Regression had an $R^2 = 0.533$ with 195 degrees of freedom

Conclusions

A quasi-experimental design was used for this project. The sections (041) and (046) were randomly assigned to using TopHat™ or not using TopHat™. The design of randomizing sections was not the same as randomizing individual students. However, it was effective for our purposes. To compensate for the lack of randomization, the two sections used control measures (e.g, SAT/ACT scores, Diagnostic) to establish a baseline. The two sections were reasonably balanced groups with no significant differences in prior knowledge. Randomizing the sections turned out to be a reasonably effective proxy for randomizing the individual students. The experimental design was set up well to evaluate the interactive, question driven, teaching method using an ARS, Tophat™, as an influential factor for improving the students' performance.

Based on the data we collected, there was an overall difference in course grades between the sections. This difference was statistically significant in *t-test* and *regression* testing. The effect of the interactive, question driven, teaching method using Tophat™ was attenuated somewhat in multivariable models that adjusted for prior knowledge, but it nonetheless remained significant. Even while adjusting for prior knowledge, the interactive, question driven, teaching method using Tophat™ appeared to produce an additional 4.98% for the overall score — half a letter grade's difference.

Since our randomization of the sections and establishing a baseline produced balanced groups, and since we also adjusted for prior knowledge, it's reasonable to attribute this difference in scores to the method using Tophat™ as an influential factor. In this setting, the interactive, question driven method facilitated by the technology appears to be working effectively. The students in the Tophat™ section appear to have made more progress during the class than those in the section without Tophat™ resulting in higher performance scores.

Practical implications

The study contributes to the literature on whether an ARS can be used to conduct a question-driven method aimed to improve student overall outcomes. The goal to help students learn in large lecture classes was achieved by using a question driven model facilitated by an ARS. On a practical, classroom management level, the results suggest that the implementation of the question-driven method with the ARS technology saved valuable time needed to cover core topics while still creating space for student-centered activities and reach to the majority of students.

Theoretical implications

We attribute the results favoring the question driven model using an ARS for a large format introductory course in mathematics because: (a) the instructional material was delivered in smaller sections interleaved with problem solving questions delivered after each part. Challenging questions were broken into smaller questions building to the whole problem. This allowed students to more easily process the instructional message and immediately apply it before moving onto the next concept or step, (b) it facilitated a uniform performance-feedback-adjustment loop for *all* students promoting self-regulation and metacognition for gauging how well they understood the lecture material and how to answer similar questions on future exams, (c) it promoted more participation because of the privacy feature afforded by the ARS (d) it promoted more attentive listening and note taking during instruction because they knew a question would soon appear, (e) all students received the same explanations for incorrect responses allowing for more uniform instruction across the class, and (f) it reduced classroom management issues that may otherwise impair learning.

Future studies may aim to provide insights on differences in students' sense of learning processes. Likewise, future studies may investigate which features of the question-based model with ARS improved learning most and why.

Methodological implications

The present study attempted to evaluate educational technology by focusing on the method afforded by the technology as opposed to the technology itself. It conducted the quasi-experimental investigation in an authentic classroom to improve ecological and methodological validity to reduce as many confounds as possible. While we controlled for prior knowledge difference between the sections other factors may have contributed to the differences. Since we conducted this study in a live course as opposed to a controlled experimental environment differences between the sections may be attributed to other factors beyond the use of an ARS such as class time schedules, quality of teaching assistants, lecture hall design differences, and other unknown factors. To minimize bias on the part of the instructor, an internal research team conducted the study. Despite these possible confounding factors researchers of this study feel confident that the results are predominantly attributed to the ARS given the rationale in the literature review of this paper. The results contribute to the pedagogical value of an ARS to facilitate instructional methods in large lecture classes.

Limitations and Future Research

In this section, we consider possible limitations of the study. The two sections differed in certain dimensions. One dimension is the medium, the other method. Our intentions were to compare the question-driven method facilitated by an ARS to the current teaching practice that massed questions into one class period without technology. We recognize that we could not have changed the current teaching practice by interleaving lecture with paper based questions and still confidently cover core course content without delays foreseen by paper administration. In terms of medium, we did not introduce the ARS into the control group because the original goal of the massed problem-solving session was to allow students time to practice at their own pace and receive as needed personalized instruction.

The researchers of this study encourage future research to experiment with audience response systems to enhance other kinds of active learning strategies aimed to improve outcomes for large enrollment foundational courses. Given the various interactive strategies there is more needed research to evaluate these methods enhanced with an ARS. For example, a future study expanding this study may investigate the use of a think-pair-share strategy along with an ARS compared to other strategies with an ARS. In conclusion, the ARS enabled a more manageable classroom to deliver problem solving opportunities than did the paper-based classroom. Pinpointing the locus of the obtained effect should be a future goal.

Lastly, future meta-analyses on technology should organize research by the method approach facilitated by the technology and not uniquely comparisons of use or not use of technology.

Implications for Practice & Policy

It is important to focus on identifying the educational challenges and goals first and then look to see how a method or strategy can help address them. A technology can help facilitate a method. Educational theory should take precedence in the process as opposed to ad hoc approaches. Instructors should look for partnerships with other instructors or educational designers to assist in course development as it is an iterative process.

Acknowledgements

A special thank you to the educational media designers Paula Torres and Carmen Louie for their hard work transforming this course from a traditional to flipped format and preparing materials for this study. They worked tirelessly to create the interactive video modules and stitch the course together. Lastly, a thank you to Dr. Jan Plass, Dr. Jennifer Hill, Dr. Ben Maddox and NYU IT leadership, and NYU's Provost for their support to not

only enhance teaching and learning with pedagogically effective uses of technology and media but also to support these efforts with research to validate them.

APPENDIX

Table 1
Study Feature Comparison for Each Section

Feature	**Description**	**Section 041 (ARS -TopHat™)**	**Section 046 (Non-ARS)**
Delayed Testing	*Was the learning outcome assessed by a delayed post-test?* (Chein et al, 2016, p. 9)	✓ Exams &Course Grades	✓ Exams & Course Grades
Baseline Control	*Were prior differences between groups (i.e., prior knowledge) were controlled by randomly assigning subjects to groups, systematically assigning subjects to groups, or using pre-test scores as covariates?* (Chein et al, 2016, p. 9)	✓ SAT/ACT scores/ GPA / Diagnostic	✓ SAT/ACT scores/ GPA / Diagnostic
One-shot	*Did the instruction sustain only one-lecture long (typically around 50 min)?* (Chein et al, 2016, p. 9)	No Instructor sustained several lecture-long classes over entire semester so as to reduce novelty effect	No Instructor sustained several lecture-long classes over entire semester so as to reduce novelty effect
Equivalent Exposure	*Did the control group receive the same in-class questions as did the experimental group?* (Chein et al, 2016, p. 9)	✓ Same questions but using a ***question-driven*** approach ***interleaved*** with the lecture-demonstrations over 2 class periods	✓ Same questions but ***question-driven*** approach ***massed*** with questions delivered in 1 class period
Repeat Questions	*Did the post-test simply repeat in-class questions?* (Chein et al, 2016, p. 9)	No	No
Peer Discussion	*Was the experimental group allowed to discuss with peers while answering in-class questions?* (Chein et al, 2016, p. 9)	✓ Both sections were allowed to work with peers or individually	✓ Both sections were allowed to work with peers or individually
Display	*Were the ARS results shown to subjects during instruction?* (Chein et al, 2016, p. 9)	✓ Results were shown to subjects during class as well as final homework results shown at end of week	Guidance and results provided during class by instructor or TA to students when requested as well as final homework results shown at end of week
Equivalent Immediate Feedback	*Did the students in both conditions receive the same immediate feedback?*	✓ Immediate Feedback was uniform for all students working on homework problems in class. Students received final scores on homework problems at the end of the week.	Students received as needed personalized guidance and instruction while working on homework problems in class. Students received final scores on homework problems at the end of the week.

Feature	**Description**	**Section 041 (ARS -TopHat™)**	**Section 046 (Non-ARS)**
Elaboration	*Were subjects provided with explanations of the correct/ incorrect answers for in-class questions?* (Chein et al, 2016, p. 9)	✓ Instructor immediately performed worked out examples/ explanations based on results of student responses to homework related problem-solving questions. Immediate feedback with explanation on performance helps students make immediate learning adjustments before moving onto the next concept.	Explanations and guidance were only provided upon request by students to TA's or instructor during the class period. Students received written explanations to homework problems at the end of the week provided by TA's.
Instructor	*Did the same instructor facilitate all conditions?*	✓	✓
Time of Day	*Did the same instructor teach all conditions during the same time of day?*	Afternoon	Morning
Points	*Did both conditions receive the same points for attendance and participation?*	✓ Attendance data collected upon entry and exit. Even though participation could be measured with the software there were no additional points attributed for answering questions. The software collected and matched data on each student's participation to their university net id. Points were only attributed to attendance. Motivation to participate was driven by use of homework problems due at the end of the week.	✓ Attendance data collected upon entry and exit. Participation in class could not be measured. Points were only attributed to attendance. Motivation to participate was driven by use of homework problems due at the end of the week.

Table 3B

Summary of Prior Ability by Gender and Section

		TopHat™ Section (041) (*n= 128*)	Section (046) (*n= 103*)
Indicators	*Measure*	*Mean (N, SD)* *t-value, p-value*	*Mean (N, SD)* *t-value, p-value*
Student prior ability	SAT/ACT Math	F = 724.9 (60, 60.4) M = 734.5 (58, 51.8) *t* = -0.92, 0.357	F = 718.6 (54, 73) M = 721.1 (40, 59.8) *t* = -0.18, 0.855
	GPA	F = 3.4 (61, 0.5) M = 3,4 (61,0.5) *t* = -0.01, 0.992	F = 3.4 (57, 0.5) M = 3.3 (42,0.5) *t* = -0.64, 0.527
	Pretest Diag	F = 77.4 (54, 11) M = 75 (57,17.9) *t* = 0.86, 0.393	F = 75.1 (53, 15.2) M = 75.6 (42,15.1) *t* = -0.16, 0.871

Image 1

Sample Homework Problem on Paper

7. The hypotenuse of a right triangle has one end at the origin and one end on the curve $y = x^2e^{-3x}$, with $x \geq 0$. One of the other two sides is on the x-axis, the other side is parallel to the y-axis. Find the maximum area of such a triangle. At what x-value does it occur?

Image 2

Sample Homework Problem Displayed in Parts Using TopHat™

HW Problem 7. The hypotenuse of a right triangle has one end at the origin and one end on the curve $y = x^2e^{-3x}$, with $x \leq 0$. One of the other two sides is on the x-axis, the other side is parallel to the y-axis. Your task is to find the maximum area of such a triangle. What is the area function you are trying to maximize?

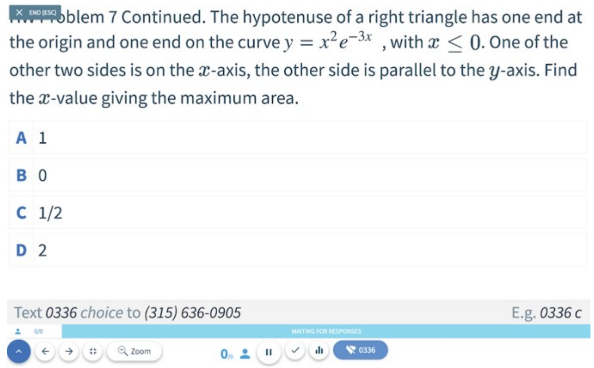
A $A(x) = \frac{1}{2}x^2e^{-3x}$

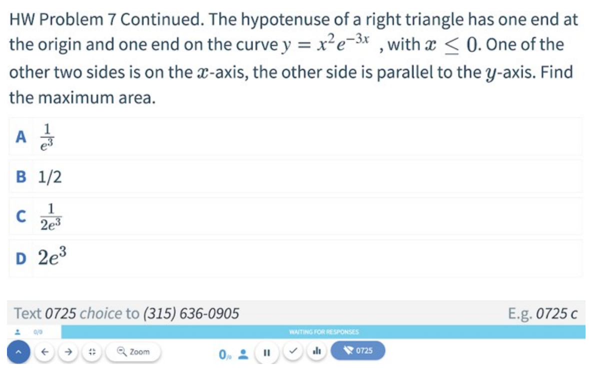
B $A(x) = \frac{1}{2}x^3e^{-3x}$

C $A(x) = x^2e^{-3x}$

Text *7086 choice* to (315) 636-0905 E.g. *7086 c*

Zoom 64 7086





References

Angelo, T. A., & Cross, K. P. (1993). Classroom assessment techniques: A handbook for college teachers. *San Francisco: Jossey-Bas*.

Azevedo, R., & Aleven, V. (2013). *International handbook of metacognition and learning technologies* (Vol. 26). Amsterdam, The Netherlands: Springer.

Banks, D. A. (2003). Using keypad-based group process support systems to facilitate student reflection. ASCILITE.

Bullock, D. W., Labella, V. P., Clingan, T., Ding, Z., Stewart, G., & Thibado, P. M. (2002). Enhancing the Student- Instructor Interaction Frequency. *The Physics Teacher,40*(9), 535-541. doi:10.1119/1.1534821

Burnstein, R. A., & Lederman, L. M. (2001). Using wireless keypads in lecture classes. *The Physics Teacher*, *39*(1), 8-11.

Butler, M., Pyzdrowski, L., Walker, V. L., & Yoho, S. (2010). Studying personal response systems in a college algebra course. *Investigations in Mathematics Learning*, *2*(2), 1-18.

Cain, J., & Robinson, E. (2008). A primer on audience response systems: current applications and future considerations. *American journal of pharmaceutical education*, *72*(4), 77.

Cardoso, W. (2011). Learning a foreign language with a learner response system: The students' perspective. *Computer Assisted Language Learn- ing*, *24*(5), 393-417.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life sciences education*, *6*(1), 9-20.

Chien, Y. T., Chang, Y. H., & Chang, C. Y. (2016). Do we click in the right way? A meta-analytic review of clicker-integrated instruction. *Educational Research Review*, *17*, 1-18.

Daniel, T., & Tivener, K. (2016). Effects of Sharing Clickers in an Active Learning Environment. *Journal of Educational Technology & Society*, *19*(3).

Dhaliwal, H. K., Allen, M., Kang, J., Bates, C., & Hodge, T. (2015). The effect of using an audience response system on learning, motivation and information retention in the orthodontic teaching of undergraduate dental students: a cross-over trial. *Journal of orthodontics*, *42*(2), 123-135.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of computer assisted learning*, *20*(2), 81-94.

D'Inverno, R. Davis, H., & White, S. (2003). Using a personal response system for promoting student interaction. *Teaching Mathematics and Its Applications, 22*(4), 163-169.

Duncan, D. (2005). *Clickers in the Classroom*, San Francisco: Pearson.

Dunn, P. K., Richardson, A., Oprescu, F., & McDonald, C. (2013). 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*, *44*(8), 1160-1174.

Ediger, Marlow. (2001). Learning opportunities in the higher education curriculum. *College Student Journal*, vol 35, no. 3, p. 410+. *Academic OneFile*, Accessed 20 June 2018.

El-Rady, J. (2006). To click or not to click: That's the question. *Innovate: Journal of online education*, *2*(4).

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

Grunert, J. (1997). The course syllabus. *A Learning-Centered Approach. Bolton: Anker*.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Hunsu, N. J., Adesope, O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers & Education*, *94*, 102-119.

Jackson, M., Ganger, A. C., Bridge, P. D., & Ginsburg, K. (2005). Wireless handheld computers in the undergraduate medical curriculum. *Medical Education Online*, *10*(1), 4386.

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education*, *53*(3), 819-827.

Liu, W. C., & Stengel, D. N. (2011). Improving student retention and performance in quantitative courses using clickers. *International Journal for Technology in Mathematics Education*. Volume 18 No 1, 51-58.

Maheady, L., & Gard, J. (2010). Classwide peer tutoring: Practice, theory, research, and personal narrative. *Intervention in School and Clinic*, *46*(2), 71-78.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. *British journal of educational psychology*, *46*(1), 4-11.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R. E. (2008). *Multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of educational psychology*, *93*(2), 390.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., ... & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary educational psychology*, *34*(1), 51-57.

Mayer, R.E., & Wittrock, M.C. (2006). Problem solving. *In Handbook of educational psychology*, eds. P.A. Alexander and P.H. Winne, 2nd ed., 287–304. Mahwah, NJ: Lawrence Erlbaum.

Mazur, E. (1997). *Peer instruction* (pp. 9-18). Upper Saddle River, NJ: Prentice Hall.

Meltzer, D. E., & Manivannan, K. (2002). Transforming the lecture-hall environment: The fully interactive physics lecture. *American Journal of Physics*, *70*(6), 639-654.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in physiology education*, *30*(4), 159-167.

Morais, A., Barragués, J. I., & Guisasola, J. (2015). Using a Classroom Response System for promoting interaction to teaching Mathematics to large groups of undergraduate students. *Journal of Computers in Mathematics and Science Teaching*, *34*(3), 249-271.

Murry, J. P., & Murry, J. I. (1992) How do I lecture thee? *College Teaching*, 40(3), 109–113.

Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, *93*(3), 223-231.

Renkl, A., & Atkinson, R. K. (2007). Interactive learning environments: Contemporary issues and trends. An introduction to the Special Issue. *Educational Psychology Review*, *19*, 235–238.

Schackow, T. E., M. Chavez, L. Loya, and M. Friedman (2004). Audience response system: Effect on learning in family medicine residents. *Family Medicine* 36 (7): 496-504.

Schoenfeld, A. H. (2002). A highly interactive discourse structure. In *Social constructivist teaching: Affordances and constraints* (pp. 131-169). Emerald Group Publishing Limited.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and instruction*, *16*(4), 475-5223.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189.

Siau, K., Sheng, H., & Nah, F. H. (2006). Use of a classroom response system to enhance classroom interactivity. *IEEE Transactions on Education*, *49*(3), 398-403.

Skuballa, I. T., Dammert, A., & Renkl, A. (2018). Two kinds of meaningful multimedia learning: Is cognitive activity alone as good as combined behavioral and cognitive activity? *Learning and Instruction*, *54*, 35-46.

Slain, D., M. Abate, B. M. Hodges, M. K. Stamatakis, and S. Wolak (2004). An interactive response system to promote active learning in the doctor of pharmacy curriculum. *American Journal of Pharmaceutical Education* 68 (5): 1-9.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary educational psychology*, *21*(1), 43-69.

Stewart, S., & Stewart, W. (2013). Taking clickers to the next level: a contingent teaching model. *International Journal of Mathematical Education in Science and Technology*, *44*(8), 1093-1106.

Terenzini, P. T., Cabrera, A. F., Colbeck, C. L., Parente, J. M., & Bjorklund, S. A. (2001). Collaborative learning vs. lecture/discussion: Students' reported learning gains. *Journal of Engineering Education*, *90*(1), 123-130.

Tivener, K. A., & Hetzler, T. (2015). The effects of an electronic audience response system on athletic training student knowledge and interactivity. *Athletic Training Education Journal*, *10*(3), 212-218.

Vrugt, A., & Oort, F. J. (2008). Metacognition, achievement goals, study strategies and academic achievement. Metacognition and Learning, 30, 123–146.

Wittrock, M. C. (1990). Generative learning processes of the brain. *Educational Psychologist*, *27*(4), 531-541.

Yuretich, R.F., Khan, S.A., Leckie, R.M., and Clement, J.J., (2001). Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course, *Journal of Geoscience Education*, v. 49, p. 11-19.