The Effects of Electronic Response Systems on Student Learning

James D. Morrison

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Industrial and Organizational Psychology Commons, Social Psychology Commons, and the Work, Economy and Organizations Commons

Recommended Citation

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.
THE EFFECTS OF ELECTRONIC RESPONSE SYSTEMS ON STUDENT LEARNING

James D. Morrison, M.A.
Western Michigan University, 2014

One of the most technologically advanced methods of implementing active student responding is the electronic response system (Judson & Sawada, 2002). This technology is known under several names including audience response system, classroom response system, and colloquially as clickers (Smith, Shon, & Santiago, 2011). To accurately assess the impact of clickers on learning performance and classroom achievement, more quantitative analysis and systematic replication of studies was needed (Kay & LeSage, 2009). This study examined the effects of ASR questions on exam performance in two sections of an organizational psychology class for majors and non-majors. A social validity questionnaire was also administered to assess student perceptions of using clickers and whether the ASR questions helped them prepare for exams. The results of the study showed no difference in performance between the two conditions. The questionnaire found that most students did not feel that the ASR questions helped them perform better on exams but that most students felt more engaged when in the ASR condition.
THE EFFECTS OF ELECTRONIC RESPONSE SYSTEMS ON STUDENT LEARNING

by

James D. Morrison

A thesis submitted to the Graduate College in partial fulfillment of the requirements for the degree Master of Arts
Industrial and Organizational Psychology
Western Michigan University
May 2015

Thesis Committee
Heather M. McGee, Ph.D., Chair
Alyce M. Dickinson, Ph.D.
Douglas A. Johnson, Ph.D.
ACKNOWLEDGEMENTS

I would first like to thank my parents for putting me in the position to be so successful by supporting me throughout my undergraduate career and encouraging me to achieve to the best of my ability. I would like to also thank my advisor, Dr. Heather McGee, for providing the support, advice, and contingencies necessary to successfully conduct this study. I am also very grateful for the analytical advice provided by Dr. Brad Huitema, without whom my statistical analysis would have been woefully incomplete. I would also like to express gratitude to my other committee members, Dr. Alyce Dickinson and Dr. Doug Johnson, for their help in increasing the precision and scholarliness of my writing and verbal behavior. Their attention to detail has helped shape and elevate this thesis to a higher academic level than I could have achieved on my own. Finally, I would like to thank Dan Sundberg, Krista Hinz, and Austin Seabert for their support while conducting this study; Dan and Krista for facilitating the study in their class and for Austin’s help with data collection and IOA.

Sincerely,

James D. Morrison
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Lecture Based Instruction

Active Responding

Choral Responding

Note Taking and Guided Notes

Response Cards

Response Systems

Research on Clickers

METHOD

Participants and Setting

Informed Consent

Materials

Experimental Design

Procedures

Classroom Attendance

Dependent Variables
# Table of Contents - Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement</td>
<td>24</td>
</tr>
<tr>
<td>Inter-observer Agreement</td>
<td>24</td>
</tr>
<tr>
<td>Fidelity of the Independent Variable</td>
<td>25</td>
</tr>
<tr>
<td>Social Validity Measures</td>
<td>25</td>
</tr>
<tr>
<td>Missing Data</td>
<td>26</td>
</tr>
<tr>
<td>RESULTS</td>
<td>28</td>
</tr>
<tr>
<td>Exam and Isomorphic Analysis</td>
<td>28</td>
</tr>
<tr>
<td>Social Validity Analysis</td>
<td>32</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>34</td>
</tr>
<tr>
<td>Overall Exam Performance</td>
<td>34</td>
</tr>
<tr>
<td>Isomorphic Question Performance</td>
<td>35</td>
</tr>
<tr>
<td>Social Validity Questionnaire</td>
<td>36</td>
</tr>
<tr>
<td>Limitations and Future Research</td>
<td>37</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>40</td>
</tr>
<tr>
<td>APPENDICES</td>
<td></td>
</tr>
<tr>
<td>TopHat Sample Screenshots</td>
<td>45</td>
</tr>
<tr>
<td>Informed Consent</td>
<td>47</td>
</tr>
<tr>
<td>Exam Coding Sheet</td>
<td>50</td>
</tr>
<tr>
<td>Anonymous Demographics and Satisfaction Survey</td>
<td>52</td>
</tr>
<tr>
<td>HSIRB Project Approval</td>
<td>56</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

1. Condition Sequence..................................................................................................................22
2. Average Unit Exam Score.........................................................................................................28
3. Average Unit Percentage of Correct Isomorphic Questions..................................................30
4. Demographic Survey Results..................................................................................................32
5. Satisfaction Survey Results.....................................................................................................33
6. Attendance Percentage per Unit Lectures...............................................................................33
LIST OF FIGURES

1. Average Unit Exam Performance by Section.........................................................29
2. Condition Interaction Plot by Section.................................................................30
3. Average Unit Isomorphic Question Performance by Section.............................30
4. Average Unit Exam Performance by Condition..................................................31
5. Average Unit Isomorphic Question Performance by Condition..........................32
INTRODUCTION

In 2010, enrollment in post-secondary education had reached 20 million people (U.S. Department of Education, 2012). This represents an overall increase of nearly 6 million students or 40% since the start of the new century. In their review of the literature, Lammers and Murphy (2002) found that a lecture format was the predominant mode of instruction in college education. Their review concluded that faculty spends 80% of class time lecturing and that 78% of faculty reported conducting lectures as their primary mode of instruction. With such a large proportion of faculty conducting lectures for such large portions of class time, it is likely that most college students receive the majority of their instruction in class from strictly lecture based delivery of content.

Lecture Based Instruction

Marr, Plath, Wakeley, and Wilkins (1960) studied the differences in academic performance of students given lecture-based instruction and self-directed learning in an introductory psychology class. The experimental group attended a class taught in a lecture format while reading the assigned textbook independently. The control group was only assigned the textbook for independent study. Lecture-based instruction was determined to be more effective than when students read the text alone. While this evidence supports the lecture format as a valid form of instruction, it offers no assessment for what parts of lecturing are effective at increasing student performance or if other methods of instruction may be more effective. Lectures could be more effective because of the examples and non-examples presented by the instructor that are not found in the text. It is also possible that students merely have poor self-management skills and are unable to allocate enough time to study appropriately (Michael, 1991). It could then be
assumed that attending lectures acts as a more effective prompt for study behaviors than giving students a list of exam dates and required reading materials. Also, lectures may provide students with clues about what will be on the exam through what material the instructor covers and what he or she emphasizes during the lecture.

According to Fredrick and Hummel (2004) there are many methods of effective instruction that traditional lecture formats do not employ. Fredrick and Hummel provide the following definition of effective instruction:

Effective instruction begins with clearly stated behavior objectives; provides accurate, competent models; provides many opportunities for active responding; delivers immediate feedback about the accuracy of responses; allows self pacing; teaches to mastery; reinforces accurate responding; and frequently and directly measures responding that is explicitly tied to the behavioral objectives, using the outcomes of those measurements to make instructional decisions. (p.11)

If instructors design their curriculum to include clear learning objectives, extensive examples and non-examples, active responding, frequent measurement of objectives, and feedback for responding they can expect to see higher learning gains over a class lacking these teaching methods.

The current study focuses on the merits of active student responding within a classroom built around these other effective teaching methods. The logic of active student responding can be found in the generative theory of learning as well. This theory states that “learners must actively engage in cognitive processing during learning by attending to relevant material, mentally organizing the selected material, and integrating the organized material with prior knowledge” (Powell, Straub, Rodriguez, & VanHorn, 2011,
A behavioral interpretation of this theory is that learners must actively verbalize information either overtly or covertly during learning, by attending to the instruction, tacting the information, and evaluating the information as it pertains to their learning history through intraverbal behavior. The idea of requiring active student responding has permeated much of the research on effective instruction ever since John Dewey (1916) posited that students learn by doing. The logic behind giving students the opportunity to engage in active responding to foster learning can be explained by some of the effects seen in the Behavior-Based Safety literature. When individuals are told how to behave safely their behavior does not change reliably (Alvero & Austin, 2004). However when those same individuals actively monitor the safe performance of others their own safe behaviors do reliably improve. So it is possible that a student that is told how to apply a concept may not be able to reliably do so later without actively engaging in the application of that concept. Therefore active responding offers the instructor the opportunity to get students engaged in applying the information presented in lecture.

**Active Responding**

Central to the idea of active responding is the learning trial, which is composed of an instructional antecedent, a student response, and teacher feedback (Randolph, 2007). There is a correlation between the number of learning trials completed and achievement; the more learning trials completed generally the better the achievement (Heward et al., 1996). This suggests that if instructors want to improve performance through effective teaching methods they must attempt to get students to actively participate several times throughout their lecture. Given the time constraints in the average college classroom, it is doubtful that students will be able to gain fluency on every topic covered during lecture.
To gain fluency students must engage in the material outside of the classroom. Active student responding within a lecture may serve other functions such as keeping student attention throughout the lecture as well as providing them with clues as to what material may be on exams or quizzes. A lecture may present material on several different topics but those topic integrated into active student responding may seem more likely to be examined from a student perspective.

One of the most common forms of active student responding is the answering of questions posed by the instructor during lecture (Lantz, 2010). Traditionally the teacher will ask a question and prompt students to raise their hand in order to respond. A limitation of this traditional approach is that it only allows for one student to be called on and respond to any given question (Heward et al., 1996). Hand-raising active responding rates are therefore limited by the number of questions asked and the number of students in each class. For large introductory classes, where there may hundreds of students in a class, it would be impractical to use this method of active student responding. Therefore, it may be more practical and effective to use a method of active student responding known as choral responding.

Choral Responding

To employ choral responding the instructor presents an instructional antecedent or prompt with a discriminative stimulus that signals the class to orally respond in unison. The discriminative stimulus is necessary to ensure that all students respond at the same time while also cueing the end of the instructional antecedent. Research in K-12 school settings has found choral responding can decrease off-task behavior, increase on-task behavior, and increase levels of correct responding (Godfrey, Grisham-Brown,
Hemmeter, & Schuster, 2003). Based on their review of the choral responding literature Blackwell and McLaughlin (2005) recommended that, because of its overall effectiveness, choral responding should be implemented by both schools and educators; citing its ability to increase on-task behavior and decrease disruptions as two of its biggest assets. Though there have been several studies demonstrating the effectiveness of choral responding in K-12 educational settings, there is limited research on the use of choral responding in a college setting, though the method of implementation and the principles that underlie its effectiveness still apply (Austin, 2000). While choral responding does not require any extra materials to use effectively, it does however pose some logistical problems (Austin, 2000). There is no way for the instructor to know what percentage of students has responded or responded correctly, especially in a large lecture environment. With a lack of research in the college setting, academics are left to speculate about its feasibility as an effective active student response method in a large lecture environment.

**Note Taking and Guided Notes**

Note taking is another practical way for all students to engage in active responding. In a study conducted by Baker and Lombardi (1985) it was found that the amount of notes and quality of information students recorded was correlated with test performance. The authors collected the notes of 94 students in an introductory psychology class for one lecture. They randomly sampled 40 of those notes for analysis. The students were likely to include main ideas but would leave out ideas they could infer from those main ideas. The students who left out information, to be inferred later, were more likely to receive lower test scores. The research into note taking is replete with
information showing this correlation between quantity and quality of notes and test
performance (Austin, 2000). It is unfortunate, in light of these findings, that skillful note
taking is absent from most college students' repertoires and that the skills are rarely
explicitly taught (Baker & Lombardi, 1985).

To ameliorate this skill deficit and increase active responding, guided notes have
been employed to provide students with the opportunity to take good notes (Austin,
2000). Guided notes can be electronic documents or paper handouts that have key
information missing from a general outline of the lecture material. With the use of
PowerPoint® in many lectures this can mean having information from slides printed on
handouts with missing words or sentences for students to complete. The learning trial
created when using guided notes is somewhat different from other methods of active
responding such as response cards or electronic response systems. The antecedents for
responding and opportunities to provide responses to those antecedents are always
available during the lecture. The “teacher” feedback to those responses is also self-
mediated; after responding the student must engage in intraverbal behavior to check the
correctness of the response as it compares to either the verbal behavior or media
presented by the instructor. The guided notes also do not improve the student's note
taking ability; to improve note taking skills they must be explicitly taught.

While support for improved student performance when implementing guided
notes is mixed, it has been found that students prefer guided notes to taking personal
notes (Neef, McCord, & Ferreri, 2006). It is more than likely that students prefer guided
notes to taking traditional notes because of the decreased response effort required during
class when using guided notes as opposed to constantly writing traditional notes.
Depending on the missing material from the guided notes it could also be possible that increases in performance are due to the targeted studying guided notes can provide. If the instructor designs the guided notes so that only key principles and related content is missing then the students using guided notes now have the essential material highlighted for them. It is also possible to account for the mixed results of the literature in that some classes are designed with study objectives so that having key information highlighted in guided notes has no additive effect to learning gains over taking traditional notes. Overall, the student that is furiously writing traditional notes is more actively engaged than the student who fills in a blank every other sentence. If the student does not have the skills or the attention span to take quality notes and it is also the goal of the instructor that those students should succeed in the course, the instructor must find a way to hold students’ attention and ensure they can create quality notes. Guided notes and other active student responding methods may provide that ability in a higher education setting that is filled with ill prepared students.

**Response Cards**

Another way to encourage active student responding is in the use of response cards. Response cards can include pre-made cards with text and colors predetermined or they can be write-on response cards on which students draw and write responses to questions as they are posed; though dry erase response boards are mostly used instead of paper cards (Knapp & Descrochers, 2009). These cards can be used to enable all students to respond to questions the instructor asks so that feedback may be given based on how the class responds (Lasry, 2008). If the majority of students respond incorrectly, then the instructor can review that material further and retest the same question later.
Godfrey et al. (2003) designed a study to compare the active student response methods of hand raising, choral responding, and response cards in a preschool setting. They randomly alternated the use of hand raising, choral responding, or response cards as the required response methods during a classroom activity. The results of the study revealed that students responded more often and were on-task a higher percentage of time when using response cards over the other two methods of responding. The instructor and students also reported they preferred using the response cards over the other two methods. Some limitations of this research are that it was not conducted in a college setting and there was no evaluation of learning gains with any of the methods of responding. These results may not transfer to a college setting where choral responding and response cards may be perceived differently by both students and instructors.

In Randolph’s (2007) meta-analysis of the research on response cards and their effects on achievement, response cards were found to increase test and quiz scores significantly in all cases when compared to hand-raising. The number of students receiving a grade of 80% or better rose from near 30% to around 62% of students when using response cards instead of hand-raising. Active participation was found to increase by nearly 50% when utilizing response cards; resulting in a large increase in learning trials completed. However, Lasry (2008) pointed out that in large classes the instructor would have to go through the tedious procedure of counting all of the responses to determine appropriate feedback. In the case of large introductory classes, response cards may pose the same logistical problems with respect to effective implementation that choral responding does.
**Response Systems**

One of the most technologically advanced methods of implementing active student responding is the electronic response system (ERS) (Judson & Sawada, 2002). This technology is known under several names including audience response system, classroom response system, and colloquially as clickers (Smith, Shon, & Santiago, 2011). Clickers have been used as a way to increase active participation in large university classrooms since the 1960s (Judson & Sawada, 2002). According to Judson and Sawada, while early use during the 1960s and 1970s focused on providing the instructor with information on student comprehension, more recently the use of clickers has shifted to conceptual questions made to start discussion. These questions still include the components of a learning trial; the antecedent, student response, and teacher feedback. However, the student response is usually in the discussion between students and the feedback is given as a general response to the topic at the conclusion of the student discussion; thus it could be argued that in many cases clickers are not being used to perform learning trials.

While the early clickers were fairly primitive mechanical units built into classrooms, today’s units are highly sophisticated three component devices: a computer program, a wireless transmitter, and a receiver (Flosason, 2010). The responses from a transmitter are taken by the receiver and interpreted with the computer program to display results instantly; this instant feedback to the instructor can be used to alter the instructor's lecture to the needs of the students.

Fallon and Forrest (2011) compared response cards to ERS and their effects on test performance. Thirty-two students from a general psychology class participated in
four review sessions held the night before each exam throughout the semester. The students were split into two groups. Each group alternated between using either an ERS or response cards to answer ten multiple choice questions during each review session. Group A used response cards in review session 1 and 3 and the ERS in session 2 and 4 while group B received the opposite ordering of conditions. No differences were found between the two methods of active responding on test achievement. While there was no difference in test achievement between the conditions, this study did not compare either of the two active response methods to a control condition with no active student responding. These results do suggest that response cards and ERS are comparable in achievement gains.

If there is no difference in achievement gains when using response cards or an ERS then there may be some practical differences an instructor should consider when planning to incorporate an active student response method into his or her classroom. A quick look at some leading producers of ERS shows that the student cost to own a response remote can be between $38 and $66 when purchasing a new device (i>clicker, 2014; Turning Technologies LLC, 2014). However, the cost for instructors is usually carried by the institution that employs them. The institution must purchase and install the affiliated wireless receivers in the classrooms and install the software on an instructor’s computer. In contrast to your average response system remote, the average price for an 11” X 16” dry erase response board can be as cheap as $2.50 and the costs to print pre-made paper cards would be even less per student (The Markerboard People, 2014). With response cards neither the instructors nor the institutions they work at are required to purchase any materials.
Research on Clickers

Brown (1972) assessed the effects of clickers on the performance of students in a non-majors mathematics course. The students were randomly assigned to one of two sections; one was taught using a student response system to review previous material at the beginning of each lecture, while the other section received no review questions or use of the response system. Depending on the answers given during this review, the instructor would discuss and drill the students on areas where they were responding incorrectly.

Both classes received the Cattell Anxiety Questionnaire, a mathematics attitude questionnaire, and a mathematics placement test, in that order, at the beginning and end of the semester. The clicker group did not have significantly better scores on the placement test; the researcher found no effect on anxiety when instructing with clickers; and the researcher found only at the .10 level of significance that clickers affected attitude toward mathematics positively. The author did note that having a positive attitude towards mathematics was significant with regard to performance, and recommended further research into the effects of clickers on attitudes and further developing the pedagogy of clicker use.

More recently Crossgrove and Curran (2008) measured the impact of clickers on non-majors and majors level biology students. They looked into the effects of clickers on student opinion, learning, and retention of material. The study was conducted across several years, with certain sections of non-majors biology and majors genetics taught with clickers for comparison with sections not using clickers. The authors did not specify the number of clicker questions asked per each semester that clickers were used, but did indicate that for fall 2006 the clickers were used between two and eight times each class.
period. Results from their 2005 sections indicated that there was no significant difference in performance for clicker and non-clicker sections. In 2006 they looked at the difference in performance for exam questions related to content covered by clickers versus content not covered using clickers; the students performed significantly better on questions based on content covered with clickers. The researchers also found that sections taught with clickers in non-majors biology had better long term retention of material as measured by a 4 month follow-up exam; they did not observe this effect in the majors genetics section tested. Regardless of class, the sections using clickers reported very high satisfaction with using clickers. Some of the results of this study may have been confounded by the fact the researchers were employing other active student responding techniques like peer teaching in their lectures (Crossgrove & Curran, 2008).

Powell et al. (2011) evaluated the effects of clickers on academic performance, student satisfaction, and perceived financial value in an introductory psychology class. The clicker section received one clicker question for every five PowerPoint® slides while the non-clicker section received a normal PowerPoint® based lecture. Performance was measured by final grades; the clicker section had on average significantly higher final grades. Powell et al. also found that 100% of the students had positive self-reported satisfaction with using clickers but found that 35% did not think purchasing clickers was worthwhile, while 32% did find purchasing clickers worthwhile. So while all the students had fun and positive experiences using the clickers, a third of the students did not feel that this experience justified their expense. This study shows that using clicker technology as formative assessment during lecture can help guide instruction to improve overall course grades. The generality of the findings may be suspect though as the groups
may have been unequal as they were not randomly assigned; there was no control for covariate variables such as GPA and the individual student improvement cannot be obtained from the aggregated average of final grades.

While the literature supports the notion that students enjoy using clickers (Kay & LeSage, 2009) and there appear to be overall learning performance increases with their use (Judson & Sawada, 2002; Kay & LeSage, 2009), the type of research that has been conducted has limitations. In their 2009 literature review of articles published from 2000 to 2007, Kay and LeSage found that a majority of the collected data were qualitative. They were unable to perform a meta-analysis because only 10 of the 67 studies they reviewed used formal statistics and quantitative measures to evaluate data and just four of those provided reliability estimates for their data collection tools. Furthermore, they found that only 11% of the data collected by the studies they reviewed was used to assess learning performance. They also found that just over one-third of the studies they reviewed examined learning and learning performance. To accurately assess the impact of clickers on learning performance and classroom achievement, more quantitative analysis and systematic replication of studies is needed (Kay & LeSage, 2009).

Since the review conducted by Kay and LeSage (2009) there have been several studies that have tried to ameliorate the lack of assessment of the effects of ERS on learning performance. One such study was conducted by Morling, McAuliffe, Cohen, and DiLorenzo (2008) in which effects of an ERS were measured on four exams across four sections of a psychology class; two sections used the response system and two did not. The students using the response system answered five multiple choice questions at the beginning of the class based on the required reading; the other sections did not see these
questions. They found that the sections using the response system had higher average exam scores; however the sections did not receive the classroom material in the same topic order which could explain why some sections performed better on different exams. The authors reported that the four sections covered similar material but that the topic order was not always the same which could lead to the sequencing of the material having an effect on the students’ ability to learn; especially if prior material did not provide them with necessary prerequisite knowledge to learn new material.

Shaffer and Collura (2009) tested the effects of an ERS on performance on 11 exam questions related to a lecture on perceptual constancy; one section used the response system and one did not. In both sections the students were asked the same four questions at the beginning of the lecture, related to the size of familiar objects such as the diameter of a traffic light or basketball hoop. The students in one section used clickers to answer the questions, while students in the other section could raise their hands to provide a response; only three students raised their hands to provide a response. The authors noted that both sections received the same lecture and subsequent discussion after viewing and answering the questions at the beginning of class. The section that used the response system had higher average performance on the 11 related exam questions; 89.49% correct for the response system group and 81.45% for the control group. In order to control for inherent group differences, the authors also measured the performance on 11 randomly chosen questions that were related to a different lecture where the response system was not used; the performances on these questions were not significantly different between the two groups. The percentage correct for these randomly chosen questions was not reported so it is not known whether the difficulty of these questions was comparable.
to the difficulty of the questions related to the response system lecture. It is also not
known whether either of these sets of questions tested more basic factual recall
information or complex concept formation than the other. The difficulty or ease of the
control questions may have accounted for similar performances across the two sections.
For example, if the 11 control questions were very easy, both groups may have performed
equally well, even if there were inherent differences between the groups. Therefore, it
may still be possible that there were inherent differences between the groups and ERS
effects seen were not simply due to the conditions they were tested under. Even though
the authors reported having the same discussion following the active response questions
in both sections, it is possible that the quality of the discussion was improved under the
clicker condition; accounting for the difference in performance on those 11 exam
questions.

Anthis (2011) conducted a study in which two sections alternated between using
an ERS to actively answer questions and just being presented with questions and
providing responses by voluntarily hand-raising. The two sections did not use the
response system for the lectures before the first exam, section B used the response system
for lectures before exam 2, section A used the response system for lectures before exam
3, and both sections used the response system for lecture before exam 4. The number of
lectures before each exam was not reported. Both sections were asked active student
response questions between one and three times per lecture. There were no significant
differences found between the two conditions. Anthis (2011) tried to control for
differences between the two sections by analyzing the differences in student grade point
average, which were not significant. This study is limited in that they only compared two

15
exams where the response system was used alternatively between sections and only gave students one to three opportunities to actively respond during the lecture. So while there have been new attempts at assessing learning performance gains relative to ERS, they suffer from several methodological flaws.

Flosason (2010) conducted a study comparing test performance between sections of a non-majors organizational psychology class with clicker use as the independent variable. Two sections of the class were taught using clickers in the fall semester and then two sections of the class were taught in the spring semester without clickers. Two instructors taught the class, one for each section during each semester. Section 1 of the fall semester used a total of 57 clicker questions while Section 2 used 59. There were two classes of lecture conducted before each unit exam for a total of eight exams. Each class was to receive four clicker questions per lecture; however time constraints caused for the removal of three questions from Section 1 and one from Section 2. All sections were taught using the same textbook and made use of additional readings and study objectives in supplementary course pack. The sequencing of the material between sections varied only slightly, and the study objectives included questions regarding the most crucial topics of the covered material. The lecture slides used by the instructor were made available to their respective students via the university’s learning management system.

Isomorphic questions to the clicker questions were presented on each unit exam; these questions were designed to be similar to the original clicker question by approaching the same concept from a different perspective or using different background information. These isomorphic questions were written such that students would need to have an understanding of and the ability to apply concepts instead of simply recalling
facts. These same isomorphic exam questions were used on the exams for both sections of the spring semester classes. Exam performance on isomorphic questions was compared across the two semesters. To control for the difference in instructors only the sections taught by each respective instructor were compared (Instructor 1: Section 1 and Section 3, Instructor 2: Section 2 and Section 4). The sections taught by Instructor 1 showed a significant difference in performance on four of the eight units, while only one of the eight units for Instructor 2 was found to be significant.

Limitations of this study include the fact that only performance on eight exam questions was analyzed. It is unknown whether performance on the entire exam was altered because of the implementation of clickers. It is possible that improved performance on those topics covered using the clickers may have improved overall learning if some topic information readily generalized to other topics not covered with the clickers. It is also possible that students may have focused their attention on the material covered using the clickers to the detriment of the material not covered using the clickers, resulting in lowered overall performance. It was also found that attendance for sections not using clickers was far lower than for those sections using clickers. This lack of attendance in non-clicker sections may have caused the overall performance of those students to be lower than the students with higher attendance using clickers. It is also possible that the different sections were not equal as they were not randomly assigned and occurred in different semesters. Finally since the material was not always delivered in the same sequence across sections it is possible that performance was altered dependent on these different sequences of instruction.
This study focused on replicating and expanding the Flosason (2010) research by utilizing a new active student responding (ASR) technology called TopHat Monocle. This new technology allows the student to respond to ASR questions with a computer, tablet, smart phone, or cell phone. The question type is also not limited to multiple choice questions; the question types can range from fill-in-the-blank, free response, matching, and sorting. This variety in question type mimics the variety of question types students can respond to with response cards but also allows for the easy tracking of student responses afforded by ERS. Most response systems only allow for multiple-choice question responding which may not transfer well to exams that are constructed with free response question types; TopHat can accommodate these question types with ease. A systematic replication of the Flosason study using this new technology was needed to account for the limitations of the previous study and to add more quantitative data to the research field on the use of ASR questions for increasing classroom achievement. The previous limitations this study controlled for were teaching material in differing sequences across sections, attendance rate between sections, and analysis of performance not just on material covered using the ERS but total exam performance. The goal of the study was to determine the effects ASR questions have on exam performance in a psychology class for majors and non-majors.

METHOD

Participants and Setting

Undergraduates enrolled in two sections, 89 total, of an organizational psychology class were selected as participants for this study. There was no formal recruitment process as each student enrolled in the class was eligible to participate. Each class met
two times every week, with a unit exam scheduled every third class meeting for a total of eight units and exams. The classes were held in a typical college classroom that was equipped with standard tables or desks for the students as well as a whiteboard, projector, and equipment necessary to display a PowerPoint® presentation. Students' current GPA and standardized test (ACT) scores were obtained in an attempt to statistically control for differences in individual performance. Demographic survey data were collected regarding the amount of time the students delegated to other activities such as employment, other classes, or extracurricular activities as well as major and undergraduate status.

**Informed Consent**

An informed consent form was distributed to the participants during the first class of the semester as well as the last day of class. The second round of informed consent was used to capture any students that may have added the class after the first day of classes. A script was used to inform the students of the nature of the study as well as voluntary consent to be part of the research. Informed consent was necessary to obtain data regarding student’s GPA and standardized test scores, as well as to obtain permission to use data collected during the study. All students used the TopHat Monocle ASR system, regardless of participation in the study; see Appendix A for screenshots of the response system. Each student was given two copies of the informed consent document along with a manila envelope. The researcher read out loud the informed consent and instructed the students on filling out the document once the researcher had left the room. The students were informed that the course instructors would not be aware of who was participating in the study. The researcher then left the room. The participants signed the consent and sealed it into the envelope if they agreed to participate. If they did
not agree, they were to seal the unsigned document into the envelope. The students kept a copy of the informed consent for their personal use. The classroom teaching assistant then collected all sealed envelopes and turned them in to the researcher. The informed consent form is included in Appendix B.

Materials

The primary material covered in the class was performance management techniques and strategies for improving performance within organizations. Each unit covered a specific topic and was taught during two class periods. The primary sources for the instructional material were a textbook, *Performance Management: Changing behavior that changes organizational effectiveness* (Daniels & Daniels, 2004) and a course-pack containing chapters and articles on performance management. Both sections received the same supplemental course pack which, in addition to the additional readings, consisted of study objectives and online resource lists. The study objectives consisted of questions related to important material and included book and article location indicators (page and paragraph numbers) for easy information retrieval.

All lecture materials were prepared in advance, collaboratively between the two instructors of both sections, so that both instructors used the same lecture slides and notes. The lecture slides were made available through the university’s learning management system. The lecture slide content was a mix of core content as well as explanations of study objectives along with prompts to use the TopHat system to present active student response (ASR) questions. The prompts for using the TopHat system were removed on days when no ASR questions are asked. The ASR questions were dispersed throughout the lecture so that approximately one learning trial occurred every 10-15
minutes for a total of five to eight questions per class. The ASR questions were constructed to assess understanding of concepts and their applications.

The ASR questions were delivered and answered using the TopHat software package. Each instructor had a unique account with TopHat that the students registered for so that they could submit their responses during class. The instructors' laptops operated the TopHat software while using PowerPoint®, and they controlled the presentation using the TopHat presentation tool. Students purchased their subscriptions to TopHat as part of the course pack required for the class. The software tracks the students' responses and can graph those responses.

Exams consisted of multiple-choice, fill in the blank, true or false, and short answer essay questions. Isomorphic questions were created for each exam corresponding to the ASR questions asked during the two lectures before each exam. Isomorphic questions are designed to be similar to the original question by approaching the same concept from a different perspective or using different background information (Flosason, 2010). The exams were based on the study objectives and lecture material covered in the two class periods before the exam. Each exam was worth the same point total, and there were two possible make-up exams throughout the semester that were not counted toward data analysis.

Experimental Design

A multiple treatment reversal design was used to compare performance with ASR and without ASR. The ASR (A) and no ASR (B) conditions were to be given in a semi-randomized order across the two sections so that neither condition occurred for more than two consecutive units of instruction. The random ordering of conditions for Section 1 was
then counter-balanced for Section 2 so that comparisons could be made across sections to control for unit difficulty and possible sequencing effects. Through randomization the actual sequence of conditions, shown in Table 1, highly resemble a traditional alternating treatment design.

The conditions in this study were not alternated rapidly as there were two days minimum between conditions, so interference between conditions was not expected. Procedural-fidelity data were collected in each condition and phase of the study and the procedure is outlined further in the independent variable fidelity section. These data were used to assess the proper implementation of ASR questions and the constancy of all other variables. Any performance improvement or maturation over the course of the semester related to extraneous variables should have influenced both conditions equally because of the alternation between conditions and the counter-balancing between sections (Wolery, Gast, & Hammond, 2010).

Table 1. Condition Sequence

<table>
<thead>
<tr>
<th>Section</th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>Unit 5</th>
<th>Unit 6</th>
<th>Unit 7</th>
<th>Unit 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Procedures

In total four units were taught with ASR questions and four without ASR questions for each section. There were 8 units covered in the entire semester, with two class periods per unit. The class was lecture based with all of the instruction delivered by one instructor per class. On days that ASR questions were used the ASR questions were presented for approximately 60 to 75 seconds depending on the individual question and time needed to respond, after which the instructor presented the answer and showed the
students the distribution of class responses. If everyone had answered before the allotted time period was over then the instructor moved on to the answer and graph of student responses. There was a minimum of one ASR question per 10-15 minutes of lecture, after an initial assessment of the first two classes it was determined that this number of questions still allowed the instructors enough time to cover the lecture material in the 75 minutes of class time. Stopping the lecture to ask an ASR question and potential technical difficulties could have reduced the amount of material covered, to avoid this, the initial assessment conducted during the first two lectures determined that one ASR question every 10-15 minutes was appropriate. On days that ASR questions were not used there were no slides containing prompts to ask ASR questions.

**Classroom Attendance**

To keep track of classroom attendance there was a short time period where the instructor activated the attendance function of TopHat at the beginning and end of class. This attendance function displays a random four digit number the instructors could show to students; the students would then respond using the device of their choice to be marked as present. To encourage attendance there was an attendance policy in place in which if students were absent for more than two class periods during the semester they lost 3 points for each additional absence. If students missed too many lectures this could result in a loss of performance on subsequent material because of a lack of prerequisite knowledge. This attendance policy was developed to help curb excessive absences and improve the number of active participants in each section.
**Dependent Variables**

The primary dependent variable was the number of points earned on unit exams. The number of points earned on each exam was compared across sections to see if either the ASR or no ASR condition resulted in higher test performance. The other variable that was assessed was the proportion of points earned on isomorphic questions that corresponded to the ASR questions presented in class. This variable was also compared across sections. Each exam was graded by the instructor of the section the test was given in. The instructors for each section created the exam for every unit together and used the same unit exams across sections. They also created the grading rubric for each exam together.

**Measurement**

Each exam was graded according to a rubric designed by both instructors. Each instructor was responsible for grading each of his or her respective section's exams. The exams were worth 35 points apiece. The number of points earned for isomorphic questions answered correctly was converted to a percentage by dividing the number of points earned by the total number of points possible and multiplying that quotient by 100.

**Inter-observer Agreement**

All exams were graded independently by a teaching assistant and the instructor of that section using a rubric designed during the creation of the exam. Once these exams were graded they were coded by the researcher using the data sheet in Appendix C. For each unit exam inter-observer agreement was obtained for a minimum of 25% of the exams by having a research assistant code the exams independently and then comparing these data sheets to the primary researcher's data sheet. Point-by-point inter-observer
agreement was calculated by dividing the number of agreements by the sum of the agreements and disagreements and then multiplying the quotient by 100. An agreement was defined as both researchers coding an exam question in the same way, either incorrect or correct. A disagreement was defined as one researcher marking a question differently from the other researcher, such as one correct and the other incorrect. The overall average IOA was 99.8% with a range of 93% to 100%.

**Fidelity of the Independent Variable**

To ensure treatment fidelity both instructors were trained in the use of the TopHat technology by the researcher prior to the start of the experiment. The researcher trained the instructors in the correct operation and handling of the presentation tool, TopHat software, and problem troubleshooting. During the study the researcher conducted treatment probes in the four sessions the instructors used the ASR system, to ensure the proper protocol was followed in the presentation of ASR questions. The probes assessed whether the ASR questions were presented with the correct verbal prompts, the full time period of 60 to 75 seconds was given to solve the problem, and that the graphed feedback and answers to the ASR questions were delivered. For Section 1, the instructor presented the ASR questions correctly 100% of the time. For Section 2, the instructor presented the ASR questions correctly 100% of the time.

**Social Validity Measures**

Each student completed a social validity questionnaire at the end of the semester, included in Appendix D. The questionnaire assessed students’ perceptions of their preference for using TopHat and the software’s effects on learning. It included questions about the perceived engagement when using TopHat as well as whether they were able to
adequately answer questions using the system. These data are useful for determining if the procedure conducted for this study is a preferred form of instruction to general lecture without ASR questions.

**Missing Data**

With all studies involving measurement of subjects across long periods of time it is possible to have subjects with missing data. There are generally three ways to categorize the mechanisms of missing data; missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Little & Rubin, 2002). Missing data are considered MCAR if the missingness does not depend on the data values that are either observed or unobserved, MAR if the missing data values do not depend on just the unobserved values, and MNAR if they are dependent on the unobserved values. Thus for this study the missing data would be considered MCAR if the missing exam scores were completely independent of the other exams or other missing exam scores; for example, if receiving a low score on a previous exam had no effect on whether that student missed the next exam and the score they would have received on that exam did not influence whether they missed it. The missing data would be considered MAR if the missing data were only independent of the unobserved exam scores; for example, if a student missed exam four because they scored highly on exams one through three, but the score they would have obtained on exam 4 did not influence their missing the exam. Finally, missing data are MNAR if the missing exam scores are dependent on the unobserved scores; for example, if a student missed an exam because they would have received a low score by taking that exam. An ANOVA routine was carried out for both sections to compare average exam scores of those students not missing exams and those
students missing exams. The results of these tests for both sections were not significant, indicating that there was no difference in average exam scores between those students missing exams and those that had not missed any exams. Given these results and lack of systematically missing data it can be assumed the missing data were at least MAR, if not MCAR.

In order to analyze the data using a general linear model a complete data set was required. The analysis could have been carried out with only those subjects who had completed all exams, or the missing data could be imputed to provide a more complete data set. Completing the analysis with a smaller sample size by excluding all subjects with missing exams would have resulted in a loss in statistical power. Operating under the assumption that the missing data were MCAR, a conditional mean regression imputation procedure was carried out to impute all missing data values (Little & Rubin, 2002). This allowed for an analysis of the complete set of subjects without decreasing the statistical power of the analysis.
RESULTS

Exam and Isomorphic Analysis

The overall exam performance and isomorphic question performance were both analyzed using a linear regression analogue to an ANCOVA analysis to determine whether the active student response system impacted performance. The analysis tested for all of the following effects: innate differences between sections not due to the treatment, interaction between the treatment and the two sections, and relative unit difficulty on performance. It was important to account for all of these variables as they could heavily influence the interpretation of the results and any main effects. There were no significant differences in performance based on the treatment condition on either the overall exam scores or the isomorphic questions with p-values of 0.342 and 0.741 respectively. This means that the treatment condition had no effect on performance for either of the dependent variables. There were also no significant differences in performance based on the section for overall exam performance or isomorphic question performance with p-values of 0.702 and 0.425 respectively. This means that the sections were not innately different and thus had roughly equivalent performance. The average score for each unit exam is displayed in Table 2 and Figure 1.

Table 2. Average Unit Exam Score

<table>
<thead>
<tr>
<th>Section #</th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>Unit 5</th>
<th>Unit 6</th>
<th>Unit 7</th>
<th>Unit 8</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>25.87</td>
<td>30.16</td>
<td>27.84</td>
<td>30.71</td>
<td>25.34</td>
<td>27.81</td>
<td>27.99</td>
<td>29.65</td>
<td>28.17</td>
</tr>
<tr>
<td>Section 2</td>
<td>27.00</td>
<td>29.73</td>
<td>26.40</td>
<td>30.23</td>
<td>26.90</td>
<td>25.13</td>
<td>28.60</td>
<td>28.94</td>
<td>27.87</td>
</tr>
<tr>
<td>Average</td>
<td>26.44</td>
<td>29.95</td>
<td>27.12</td>
<td>30.47</td>
<td>26.12</td>
<td>26.47</td>
<td>28.30</td>
<td>29.30</td>
<td></td>
</tr>
</tbody>
</table>
After looking for overall effects the data were examined to determine if there were any differential effects due to the unit and treatment sequencing. There was a significant interaction between the intervention and the two sections for the overall exam variable, having a $p$-value of 0.022, indicating a difference in the treatment condition’s effect across the two sections. Section 2 had higher average performance than Section 1 under the ASR condition, and Section 1 had higher average performance than Section 2 under the control condition. A plot of this interaction can be seen in Figure 2. This may indicate beneficial effects of the treatment condition for Section 2, with detrimental effects in Section 1, or that there were innate differences in unit difficulty. The interaction effect for the isomorphic questions across the two class sections was not significant, having a $p$-value of 0.149. The average percentage of correctly answered isomorphic questions for each section is presented in Table 3 and in Figure 3.
Figure 2. *Condition Interaction Plot by Section*

Table 3. *Average Unit Percentage of Correct Isomorphic Questions*

<table>
<thead>
<tr>
<th>Section #</th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>Unit 5</th>
<th>Unit 6</th>
<th>Unit 7</th>
<th>Unit 8</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>86.45</td>
<td>91.74</td>
<td>94.96</td>
<td>87.21</td>
<td>72.00</td>
<td>85.07</td>
<td>84.25</td>
<td>89.08</td>
<td>86.35</td>
</tr>
<tr>
<td>Section 2</td>
<td>83.18</td>
<td>90.09</td>
<td>87.77</td>
<td>89.28</td>
<td>73.27</td>
<td>78.88</td>
<td>78.52</td>
<td>88.60</td>
<td>83.70</td>
</tr>
<tr>
<td>Average</td>
<td>84.82</td>
<td>90.92</td>
<td>91.37</td>
<td>88.25</td>
<td>72.64</td>
<td>81.98</td>
<td>81.39</td>
<td>88.84</td>
<td></td>
</tr>
</tbody>
</table>
To determine if the interaction effect was due to treatment effects or innate differences in unit exam difficulty a general linear model was fit to test the hypothesis that the unit’s exams and their respective isomorphic questions were equivalent. The test of whether the unit exams were equivalent to each other and the test of whether the isomorphic questions for each exam were equivalent, were both significant with $p$-values of 0.008 and 0.003 respectively. These outcomes indicate that the unit exams were not equivalent to each other and the isomorphic questions for each unit exam were also not equivalent to each other. This means that any difference in performance for both measures between units is due to the innate difficulty differences in the actual exams and isomorphic questions, and not to any treatment or section differences. The average performance for each unit exam by condition and average performance for isomorphic questions by condition are presented in Figure 4 and Figure 5 respectively.

![Figure 4. Average Unit Exam Performance by Condition](image)
Social Validity Analysis

In addition to the quantitative analysis of the ASR system a qualitative survey was given to student to assess both class demographics as well as satisfaction with the ASR system. Both GPA and ACT scores were validated separately using electronic student records. The demographic portion of the social validity questionnaire revealed that the students’ psychology background and extracurricular work load were very similar across the sections, see Table 4. The two sections had nearly identical average GPA, credit hour course load, and average employment hours. Their responses to the satisfaction portion of the survey were also very similar; these questions appear in Table 5.

Table 4. Demographic Survey Results

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Section 1</th>
<th>Section 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average GPA</td>
<td>3.37</td>
<td>3.27</td>
</tr>
<tr>
<td>Average ACT</td>
<td>25.88</td>
<td>21.89</td>
</tr>
<tr>
<td>Average number of psychology courses taken previously</td>
<td>4.6</td>
<td>5.5</td>
</tr>
<tr>
<td>Percentage employed during semester</td>
<td>79%</td>
<td>85%</td>
</tr>
<tr>
<td>Average hours per week working</td>
<td>22.3</td>
<td>23.9</td>
</tr>
<tr>
<td>Average credit hours being taken currently</td>
<td>13.6</td>
<td>14</td>
</tr>
</tbody>
</table>

Figure 5. *Average Unit Isomorphic Question Performance by Condition*
Only 35% and 41% of Section 1 and Section 2, respectively, said that the questions asked using TopHat helped them prepare for the exams. A large majority from both sections stated that they felt their instructor was engaging without using the TopHat system, but 79% of Section 1 and 59% of Section 2 said they felt more engaged when answering questions using the TopHat system. However, only 35% of Section 1 and 38% of Section 2 stated they would like to use the TopHat system again in another class. Data were also taken on the attendance rate of each section for each unit’s lectures (see Table 6). Again, both sections had nearly equivalent attendance rates for each unit during the semester. Section 1 attendance ranged from 78% to 98%, while Section 2 ranged from 81% to 99%.

Table 5. Satisfaction Survey Results

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>Section 1</th>
<th>Section 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used clickers before</td>
<td>82%</td>
<td>79%</td>
</tr>
<tr>
<td>Used TopHat or something similar before</td>
<td>32%</td>
<td>15%</td>
</tr>
<tr>
<td>Felt TopHat helped them prepare for exams</td>
<td>35%</td>
<td>41%</td>
</tr>
<tr>
<td>Felt they scored higher on exams with material covered using TopHat</td>
<td>29%</td>
<td>38%</td>
</tr>
<tr>
<td>Felt questions asked in class were comparable to questions on exam</td>
<td>65%</td>
<td>59%</td>
</tr>
<tr>
<td>Felt number of TopHat questions asked in class was appropriate</td>
<td>76%</td>
<td>82%</td>
</tr>
<tr>
<td>Felt amount of time provided to answer TopHat questions was appropriate</td>
<td>100%</td>
<td>79%</td>
</tr>
<tr>
<td>Felt more engaged with TopHat</td>
<td>79%</td>
<td>59%</td>
</tr>
<tr>
<td>Felt instructor was engaging without TopHat</td>
<td>94%</td>
<td>85%</td>
</tr>
<tr>
<td>Would like to use TopHat in another class</td>
<td>35%</td>
<td>38%</td>
</tr>
</tbody>
</table>

Table 6. Attendance Percentage per Unit Lectures

<table>
<thead>
<tr>
<th>Section #</th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>Unit 5</th>
<th>Unit 6</th>
<th>Unit 7</th>
<th>Unit 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1</td>
<td>98%</td>
<td>78%</td>
<td>88%</td>
<td>82%</td>
<td>93%</td>
<td>85%</td>
<td>90%</td>
<td>78%</td>
</tr>
<tr>
<td>Section 2</td>
<td>99%</td>
<td>81%</td>
<td>91%</td>
<td>94%</td>
<td>83%</td>
<td>84%</td>
<td>91%</td>
<td>88%</td>
</tr>
</tbody>
</table>
DISCUSSION

Overall Exam Performance

The goal of this study was to assess the effects of using an ASR system during lecture on exam performance. There was no significant effect of using ASR with the TopHat system on student exam performance. These results contradict the findings of many published studies (Kay & LeSage, 2009). It is possible that the ASR condition did not affect performance because there were too few opportunities presented for students to respond to, and thus there were no improvements due to a lack of practice of the material. At most, students were only presented with six different ASR questions per lecture, with 12 questions per unit. It is, however, much more likely that there were no differences in performance because the majority of learning took place outside of the classroom. Both sections were provided with study objectives and supplemental materials in addition to the general lecture outlines. These additional materials could have allowed students to perform well, regardless of their exposure to ASR questions during lecture.

There was a significant interaction effect observed between the treatment and the two sections for overall exam performance, as seen in Figure 2. This could be interpreted as the ASR condition having a positive effect on performance for Section 2 and detrimental effect on performance for Section 1. However, the significant result for the test of unit exam equivalence indicates that the units were innately different from each other. One possible interpretation of these differences would be that some tests were in general more difficult than others either through the material they covered or the relative difficulty of the questions asked on each exam. This effect appears to be observed in the differences between units 1-4 in Figure 1. It is also possible that unit 5 exam performance was
negatively impacted because of the Spring Break holiday. This break occurred between the presentation of unit 5 lectures and the unit 5 exams for both sections. This is also the period in which the classroom material switches from the textbook to material covered across journal articles. There appears to be a historical drop in performance between exam 4 and 5 because of the learning curve required to become adept at reading and analyzing journal articles. This is not surprising as most undergraduate students are very familiar with the structure of textbooks but may have read very few scientific journals during their short academic careers.

**Isomorphic Question Performance**

There was also no significant treatment effect observed for the isomorphic question performance. This lack of effect could have been an artifact of the questions selected for use as ASR questions. These questions appear to have created a ceiling effect in performance. The median performance value for every unit across both sections was equal to or larger than their respective mean performance, with the mode performance in half of the units observed equaling 100% between the two sections. It is possible that if the ASR questions had been relatively more difficult then there could have been more discernable performance increases on their isomorphic counterparts. More difficult questions would have been more likely to be answered incorrectly during the lecture, thus providing the opportunity for feedback in the form of correction and clarification. This corrective feedback would be more likely to increase the performance on isomorphic questions, so those students exposed to the ASR condition would likely have performed better.
There was also a significant effect observed when testing for the equality of the isomorphic questions, indicating that any differences between the units were due to innate differences in each unit’s isomorphic questions. However this effect appeared to be due to the large difference in performance on the unit five isomorphic questions. With an additional test of the equality of isomorphic questions, excluding unit 5 performance, the $p$-value for the test raised to 0.070 indicating that all isomorphic questions were roughly equivalent with the exception of unit 5. This result may also be confounded for the same reasons mentioned previously about the unit 5 exam; the change in readings from textbook to journal articles and the temporal interruption of the Spring Break holiday between lecture and exam delivery.

**Social Validity Questionnaire**

One of the more interesting parts of the social validity survey was that students generally disagreed with the idea of the response system aiding them in their preparation for exams. Between both sections eight students strongly disagreed, 34 disagreed, 25 agreed, and only one strongly agreed that answering questions using TopHat helped them prepare for exams. This may have been due to the study objectives and supplemental exercises provided in the class course pack. The additional questions asked during class could have been unnecessary for students to rely on in order to perform adequately on the exams.

It is surprising to note that the majority of students did not want to use TopHat again in another class, in spite of a majority of students indicating they felt more engaged when using the TopHat system to answer questions during class. Part of this effect may be an artifact of the way in which the survey was constructed; there were only options for the
students to strongly disagree, disagree, agree, or strongly agree with each statement. It is possible some students did not feel either engaged while using TopHat or during regular lecture and so defaulted to agreeing with the latter statement but would still not want to use TopHat in the future. It could also have been due to the fact that many students had been using clickers in previous classes and were opposed to spending additional money on the recurring semesterly TopHat fee for using the system. Previous research has indicated that students may enjoy response systems but are reluctant to spend money on them (Powell et al., 2011).

**Limitations and Future Research**

It is possible that there were not enough opportunities to engage in active student responding under the ASR condition in this study to have any beneficial effects on performance. While other studies, such as Crossgrove and Curran (2008), have indicated that performance can be improved with similar amounts of responding, those studies had not indicated the same level of methodological control as the current one. When testing is frequent, thus performance feedback in general is frequent, and this is combined with the provision of study objectives and supplemental study materials, it could be that active student responding provides no added benefit to student performance. Similarly constructed studies could improve upon this limitation by systematically manipulating the quantity of ASR questions to determine what amount of ASR is required to achieve significant improvements in performance. For now it may be accurate to recommend that those classes lacking study objectives or supplemental study materials should consider incorporating some form of ASR to offset this lack of class structure.
Another limitation of this study appears to be the difference in relative difficulty of unit exams. Either through the content they covered or the design of the exam questions themselves, the unit exams were not equivalent to each other. Future studies should formally investigate the equivalence of any of their dependent measures before implementation to avoid this shortcoming. Without accounting for these innate differences it is possible for someone to conclude they have found a positive or negative effect of ASR on performance when it is simply an innate difference in the dependent variables they are comparing.

While it may not be surprising that ASR had no discernable effect on overall exam performance, it was very surprising it had no effect on isomorphic question performance. Any possible effect may have been mitigated by the relative difficulty of the ASR questions and their respective isomorphic questions. In this study there appears to be a ceiling effect on performance with regards to isomorphic questions performance. Future studies should attempt to control for the relative difficulty of any questions they ask with their response systems and on their dependent measures to avoid any ceiling or floor effects in performance.

While there are some limitations to this study, the detailed and exact methodology will allow any future researchers to accurately and systematically replicate it. Much of the research in active student responding, especially in the case of clickers, has lacked much needed quantitative assessment of learning performance (Kay & LeSage, 2009). Furthermore, as pointed out in the literature review, many of the published studies have errors in methodology. These errors include limited exposure to the ASR condition (Anthis, 2011), differing administration of class content (Morling et al., 2008), and
inadequate control conditions (Shaffer & Collura, 2009). This makes suspect the results and conclusions the community is able to confidently draw from them. Systematic, quantitative, and methodologically sound research is needed to understand if students actually benefit from the use of active student responding technology.
REFERENCES

Alvero, A.M. & Austin, J. (2004). The effects of conducting behavioral observations on

Anthis, K. (2011). Is it the clicker or is it the question? Untangling the effects of student
doi:10.1177/0098628311411895

(Eds.), *Handbook of applied behavior analysis* (pp. 449-472). Reno, NV: Context
Press.

doi:10.1207/s15328023top1201_9

response cards to increase student performance. *The International Journal of
Special Education, 20*(2), 1-5.

Brown, J. D. (1972). An evaluation of the Spitz Student Response System in teaching a
course in logical and mathematical concepts. *The Journal of Experimental

biology course: Student opinion, learning, and long-term retention of course
material. *CBE-Life Sciences Education, 7*(1), 146-154. doi:10.1187/cbe.07-08-0060


APPENDIX A

TopHat Sample Screenshots
APPENDIX B

Informed Consent
You have been invited to participate in a research project. This project will serve as James Morrison’s thesis for the requirements of a master’s degree. This consent document will explain the purpose of this research project and will go over all of the time commitments, the procedures used in the study, and the risks and benefits of participating in this research project. Please read this consent form carefully and completely and please ask any questions if you need more clarification.

**What are we trying to find out in this study?**
We are examining the functionality and value of a new Active Student Response technology for use in the classroom. We would like to use data that you generate during this course as research data.

**Who can participate in this study?**
Students age 18 and up who are enrolled in PSY3440 over the spring 2014 semester are eligible for participation.

**Where will this study take place?**
The study will take place in the regularly scheduled classroom.

**What is the time commitment for participating in this study?**
There is no time commitment for participating in the study above and beyond what is expected for participation in the PSY3440 course.

**What will you be asked to do if you choose to participate in this study?**
The practices involved with the course will remain the same regardless of participation or non-participation in the study. We are requesting permission to use your answers in class and on exams as research data and your GPA as a control for comparing your data across class sections. Your instructor will not be aware of who is participating in this study and your grade will in no way be affected by your participation. All data will remain confidential.

**What are the risks of participating in this study and how will these risks be minimized?**
There are no known risks for participating in the study above and beyond normal risks associated with participation in the PSY3440 course.

**What are the benefits of participating in this study?**
There are no direct benefits to you personally; information gathered may be useful for improving general teaching practices.

**Are there any costs associated with participating in this study?**
There are no costs associated with participation in the study.

**Is there any compensation for participating in this study?**
No.
Who will have access to the information collected during this study?
The student investigator, principle investigator, and research staff will have access to the data collected in the study. The data will remain confidential and will be stored in a locked filing cabinet in a locked office of the principle investigator at Western Michigan University for four years. After four years, the data will be destroyed.

What if you want to stop participating in this study?
There are no consequences for withdrawing from the study. If you decide to discontinue with the study, you may inform the student investigator or the principle investigator. You can choose to stop participating in the study at anytime for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences academically or personally if you choose to withdraw from this study. Further, your decision will have no effect on your relationship with Western Michigan University.

Should you have any questions prior to or during the study, you can contact the primary investigator, Dr. Heather McGee at 269-387-4460 or heather.mcgee@wmich.edu. You may also contact the Chair, Human Subjects Institutional Review Board at 269-387-8293 or the Vice President for Research at 269-387-8298 if questions arise during the course of the study.

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.

I have read this informed consent document. The risks and benefits have been explained to me. I agree to take part in this study.

By signing this consent document I am giving my permission for data I provide in the course to be used as research data.

___________________________________
Please Print Your Name

___________________________________
Participant’s signature

___________________________________
Date
APPENDIX C

Exam Coding Sheet
<table>
<thead>
<tr>
<th>Questions:</th>
<th>Points Available</th>
<th>Points Earned</th>
<th>Isomorphic (yes/no)</th>
<th>Points Available</th>
<th>Points Earned</th>
<th>Isomorphic (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D

Anonymous Demographics and Satisfaction Survey
1. How many psychology courses have you taken prior to this semester?

   Answer: __________

2. How many credit hours are you taking this semester?

   Answer: __________

3. What is your undergraduate student’s status? (circle one)

   Freshman     Sophomore     Junior     Senior

4. Were you employed during the semester? (circle one)

   Yes     No

5. If so, how many hours per week did you work on average?

   Answer: __________

6. Gender (circle one)

   Male     Female

7. How old are you?

   Answer: __________

8. Have you ever used clickers (wireless response systems) during your education (K-12 or college)?

   More than 3 classes     In 2-3 classes     In one class     Never
9. Have you ever used TopHat or a similar active student response system during your education (K-12 or college)?

More than 3 classes    In 2-3 classes    In one class    Never

10. The active student response questions asked using the TopHat system helped me prepare for exams.

   Strongly Disagree    Disagree    Agree    Strongly Agree

11. I felt that I did better on the exams that tested material we had covered using the TopHat system.

   Strongly Disagree    Disagree    Agree    Strongly Agree

12. The active student response questions were comparable to the questions asked on exams.

   Strongly Disagree    Disagree    Agree    Strongly Agree

13. I feel that the number of TopHat questions per lecture (when used) was appropriate.

   Strongly Disagree    Disagree    Agree    Strongly Agree

14. I feel that the amount of time provided to answer TopHat questions (when used) was appropriate.

   Strongly Disagree    Disagree    Agree    Strongly Agree

15. I felt more engaged during class when there were active student response questions asked using the TopHat system.

   Strongly Disagree    Disagree    Agree    Strongly Agree
16. The instructor was able to keep me engaged without using the TopHat questions.

Strongly Disagree  Disagree  Agree  Strongly Agree

17. I would like to use the TopHat system again in another class.

Strongly Disagree  Disagree  Agree  Strongly Agree

Please write down any additional comments you have about the TopHat system for active student responding:
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
APPENDIX E

HSIRB Project Approval
Date: October 30, 2013

To: Heather McGee, Principal Investigator
James Morrison, Student Investigator for thesis

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number 13-10-35

This letter will serve as confirmation that your research project titled “The Effects of Electronic Response Systems of Student Learning” has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may only be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., you must request a post approval change to enroll subjects beyond the number stated in your application under “Number of subjects you want to complete the study.”) Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: October 30, 2014

251 W. Walwood Hall, Kalamazoo, MI 49008-5436
PHONE: (269) 387-8279 FAX: (269) 387-8276