EXPLORING THE EFFICACY OF ELECTRONIC RESPONSE DEVICES IN NINTH-GRADE SCIENCE CLASSROOMS

by

JOHN A. DEY JR.

A DISSERTATION

Presented to the Department of Educational Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Education

June 2009

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

John Dey

Title:

"Exploring the Efficacy of Electronic Response Devices in Ninth-Grade Science Classrooms"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Education degree in the Department of Educational Leadership by:

Gerald Tindal, Chairperson, Educational Leadership
Keith Hollenbeck, Member, Educational Leadership
Kathleen Scalise, Member, Educational Leadership
Jean Stockard, Outside Member, Planning Public Policy & Mgmt

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

June 13, 2009

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

John A. Dey Jr. for the degree of Doctor of Education

in the Department of Educational Leadership to be taken June 2009

Title: EXPLORING THE EFFICACY OF ELECTRONIC RESPONSE DEVICES IN NINTH-GRADE SCIENCE CLASSROOMS

Approved: ______________________________

Dr. Gerald Tindal

Student use of electronic response technology has been prevalent in postsecondary institutions and is beginning to penetrate K-12 classroom settings. Despite these trends, research exploring the impact of this technology in these settings has been limited. The extant research has relied heavily on survey methodologies and largely has focused on student/teacher perception or implementation practices while remaining silent on learning outcomes. The purpose of this study was to broaden the scope of research models used to explore electronic response technology and its impact on student learning. The study took place in a ninth-grade science classroom at a large high school with a comprehensive curriculum. Study participants were first-year high school students enrolled in one of two sections of the freshman science sequence focusing on Physical Science content. One section, serving as the Treatment group, used electronic response devices on a daily basis to respond to preplanned teacher questions.

The other section, serving as the Comparison group, relied on traditional methods of interaction such as raising hands to respond to questions. They responded to the same set of preplanned questions and differed only in the manner of response, with the teacher asking the class and then calling on one of the students to answer. The study focused on academic achievement, as measured by student performance on a pre- and posttest, as well as student engagement, measured by momentary time sample data taken throughout the entire class with focused attention on periods of teacher questioning. The analysis of academic achievement employed an ANOVA, and no statistically significant difference was found between the groups. Engagement data were analyzed using an independent samples *t* test, and statistically significant differences were found between the two groups. Findings from this study indicated that, when using electronic response technology in their science classes, students demonstrated significantly higher levels of engagement across an entire class period as well as during teacher questioning. Implications of the study have been framed around the promise of electronic response technology for engaging and motivating students.

CURRICULUM VITAE

NAME OF AUTHOR: John A. Dey Jr.

PLACE OF BIRTH: Beaufort, South Carolina

DATE OF BIRTH: November 21, 1971

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon
College of Charleston

DEGREES AWARDED:

Doctor of Education in Educational Leadership, 2009, University of Oregon
Master of Education in Education, 2002, University of Oregon
Bachelor of Arts in Biology, 2001, University of Oregon

AREAS OF SPECIAL INTEREST:

Instructional Technology
Organizational Assessment
Educational Methodology

PROFESSIONAL EXPERIENCE:

Science Teacher, Eugene School District, Eugene, Oregon, 2008-present

Assistant High School Principal, Springfield School District, Springfield, Oregon, 2005-2008

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Drs. Gerald Tindal and Keith Hollenbeck for their guidance and assistance over the past 7 years—not the least of which was their assistance in preparation of this manuscript. In addition, special thanks are due to Tracy Bouselott, whose curiosity and commitment to high-quality instruction were critical to the planning and completion of this investigation.

DEDICATION

This work is dedicated to all teachers who not only wish to work in public education but also strive to impact the field. In addition, I dedicate this work to my father, whose sacrifice and example made it possible for me to complete an endeavor of this magnitude.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

This study explored the role electronic response systems (ERS) play in increasing student learning in ninth grade science classrooms. While technology is prodigious throughout contemporary communication and entertainment (Friedman, 2005; Pew Internet & American Life Project [PIALP], 2008) today's students are increasingly incorporating digital technology into their academic pursuits (Corporation for Public Broadcasting [CPB], 2002), with schools and school districts following suit.

Draper (1998) offered a caution, though, about incorporating digital technology into academics when he stated that instructional technology was truly valuable only when it addressed a deficiency in traditional methods of instruction. One deficiency of traditional instructional practice was the inability to consistently engage all students in the learning process (Caldwell, 2007; Newmann, 1992; Reay, Li, & Bao, 2008)—an important point given that engagement was necessary for learning (Beatty, 2004; Carini, Kuh, & Klein, 2006; Newmann, 1992; National Research Council and the Institute of Medicine, 2004). Research suggested there were practical ways to increase student engagement, one of which was to incorporate active learning strategies into the classroom (Beatty, 2004; Chickering & Gamson, 1987; Crouch & Mazur, 2001; Freeman et al., 2007; Marrs & Novak, 2004). Active learning strategies focused on reading, writing, reflecting and/or discussing what students did as part of the learning process (Freeman et al., 2007; Smith et al., 2005; Watkins & Slocum, 2003). Used effectively, instructional technology can efficiently allow for active pedagogical strategies (Crouch & Mazur, 2001; Hake, 1998; Marrs & Novak, 2004; Novak, Patterson, Gavrin, & Christian, 1999). Instructional technology that showed promise at increasing active learning was the electronic response system (Duncan, 2005; Guthrie & Carlin, 2004; Preszler, Dawe, Shuster, & Shuster, 2007; Roschelle, Penuel, & Abrahamson, 2004).

Now, more than ever, technology influences the world in which students live and the way they interact with that world (CPB, 2002; Friedman, 2005). In contrast to previous generations, today's students have always known digital technology to be an everyday part of their lives. Consequently, students are not only familiar with digital technology, but they also seek out ways to integrate digital technologies—e.g., computers, the Internet, instant messaging and cell phones—into their lives (United States Department of Education [USDOE], 2004). Citing evidence of the emerging status of technology among modern students, CPB (2002) reported that time spent using digital media by children aged 13-17 was greater than the amount of time spent watching television. In their report on the New Digital Ecosystem, the Pew Internet & American Life Project (2008) stated that technology gadgets were ubiquitous in contemporary society. Supporting this claim was the fact that 88% of college students owned cell phones, 81% owned digital cameras, 63% owned MP3 players and 55%

owned laptop computers (PIALP, 2008). Beyond the realm of college students, in 2002, 83% of family households in America reported owning computers (PIALP, 2008). The Internet was at the center of this technological explosion. *Connected to the Future*, a report by the CPB (2002), likened the burgeoning increase in the use of the Internet to an adolescent growth spurt. Between 2000 and 2002 there was a 59% growth rate in the use of the Internet, with 78% of American children between the ages of 12-17 accessing the World Wide Web.

While the Internet was used primarily for entertainment or communication, students increasingly used digital technology in schools and in doing schoolwork. For example, 99% of U.S. college students used a computer as part of their academic endeavors, and students aged 6-17 reported educational activities such as homework or research for a paper among their top five reasons for using digital technology (CPB, 2002). According to the survey conducted by the Center for Public Broadcasting, 71% of students used digital technology (the Internet) to find resources for their last major paper, as opposed to only one quarter of the students using more traditional reference materials (PIALP, 2008).

Aside from research for papers and reports, students utilized technology in a myriad of ways to support their learning. For example, more than half of the student respondents to a survey conducted by the Pew Internet & American Life Project (2008) indicated they used a website set up by a school or a class, one third downloaded a study aid, and 17% created a web page for a school project. Computers and the Internet were not the only sources of digital technology used by students for educational reasons. Sixty percent of students in Grades 6-12 used instant messaging to contact adults (e.g., family members, teachers and coaches) on a weekly basis, and 41% of on-line teens used E-mail and instant messaging to contact teachers or classmates about schoolwork.

Although growth in the access and use of digital technology from 1990–2008 was staggering, it will likely pale in comparison to the next 10 years. Given that computing power doubles every 18 months (Moore's Law), that storage power doubles every 12 months (disk law) and that communications power doubles every 2-3 years with improvements in fiber optics and compression (Gilder's Law), it is safe to state that students lead increasingly more technological lives (PIALP, 2008). Guthrie and Carlin (2004) proposed that students were just as capable as in the past, but because they interacted with such a technology-rich life outside of school, they were waiting to interact with content in ways that were immediately interactive and provided almost instantaneous feedback. Because of the rocketing increase in use among teenagers, it is clear digital technology engages students. In Chapter II, I argue that if engagement affects the amount of time people are willing to devote to learning (Bransford, Brown and Cocking, 1999), and technology engages today's students, it is logical to assume that technology holds promise for increasing student learning.

Defining Engagement

While the research literature lacks a concise and universally accepted definition of *engagement*, researchers and practitioners accepted it as a prerequisite for learning (Carini et al., 2006; Chickering & Gamson, 1987; Newmann, 1992; Tallal, Merzenich, Jenkins, & Miller, 1999). In the subsequent review of the research, I offer a definition of engagement and described the construct in behavioral and emotional terms. Behaviors such as persistence, attention and effort served as observable exemplars of engagement (National Research Council and the Institute of Medicine, 2004). Less easily observed were emotions associated with engagement. However, pride in success served as a good proxy. With respect to the behaviors associated with engagement, cognitive behaviors were most important (Newmann, 1992). Given the role of engagement in learning, it was reasonable to accept that increases in learning would require increased levels of engagement. The next chapter's review of literature on teaching and learning reasons that active learning strategies were effective at increasing student engagement.

Active Learning

Views on what constitutes *active learning* vary. While all learning, by definition, could be considered an active process, Chickering and Gamson (1987), Guthrie and Carlin (2004), and Watkins and Slocum (2003) suggested active learning required more reading, writing, talking, listening and reflecting than what was involved in a traditional lecture approach to instruction. Despite a lack of consensus regarding a concise definition, active learning has been the subject of a significant amount of research literature (Freeman et al., 2007; Hake, 1998, 2002; Marrs & Novak, 2004; Meltzer & Manivannan, 2002; Novak et al., 1999; Smith et al., 2005; Udovic, Morris, Dickman, Postlethwait, & Wetherwax, 2002). This research supports the role of active learning in increasing student engagement. For example, Chickering and Gamson (1987) pointed to active learning as one of seven principles for increasing student engagement in learning. Guthrie and Carlin (2004) highlighted active learning strategies as more important than ever, as students, accustomed to a highly interactive world, were primarily active learners. Therefore, traditional methods of instruction were increasingly out of touch with student learning styles. Finally, Hake (1998) compared student performance in interactive-engagement methodological classes to student performance in more traditional, lecture-based, classes. Hake found that students in interactive classes performed at rates significantly higher than their peers in classrooms and lecture halls employing traditional methods of instruction. In that study, Hake differentiated interactive-engagement methods from traditional methods of instruction as those methods employing hands-on and heads-on activities. His findings supported the idea that active learning methods increased student engagement and in turn resulted in higher levels of student achievement. I proposed that one method for increasing the active engagement in a classroom was the introduction of technology.

Technology and Motivation

In order for students to actively engage in the learning process, they must be motivated to do so. McGregor (1999) defined motivation, specifically the motivation to learn, as the process of initiating, sustaining, and directing activity towards learning. Intrinsic and extrinsic factors affect the degree to which someone is motivated to learn (Jacobson & Xu, 2002). Extrinsic factors are external and most often tangible—rewards and grades being good examples. Intrinsic factors affecting motivation are intangible and related to emotions such as self-determination, satisfaction and competence (Lowman, 1990). The research on motivating factors varies (Lowman, 1990; Sotto, 1994). Although extrinsic motivators have been proven to work, they must be given indefinitely to effectively influence motivation (Lowman, 1990), undermine intrinsically motivating factors and result in poor learning (Sotto, 1994). Although intrinsically motivating factors are more ambiguous and slower to take effect, they generally have a longer lasting impact and result in higher quality learning (Sotto, 1994).

Four-Factor Theory of Motivation

For the purposes of this study, I evaluated the impact of technology on intrinsic motivation by using Keller's (1987) four-factor theory of motivation, the ARCS Motivation Model Theory. In his model, Keller posited there are four factors influencing motivation: The first was attention (A), the second was relevance (R), the third was competence (C) and the last was satisfaction (S).

Attention

Capturing the students' attention is critical to stimulate motivation in Keller's model. This idea supported the stance of Tallal et al. (1999) that students must first attend to the content in order to learn. Keller (1987) explained that a student's attention had to be captured and sustained in order to initiate motivation. Technology-enhanced methods of instruction, as compared to more traditional methods, are motivating because students' use of technology (as opposed to teachers' use or no use at all) is novel. Additionally, technology-enhanced instruction mixed with lecture and discussion adds the variability within instruction needed to maintain students' attention (Fox, 1988; Jacobson & Xu, 2002).

Relevance

Relevance is the second factor that Keller's (1987) model requires for initiating and maintaining motivation. Students must see relevance in either the content covered or, at least, the methods used to explore the content. With traditional instructional strategies, students are often passive recipients of the content and as such can easily disengage (Guthrie & Carlin, 2004), as they see no connection between the activity and their personal goals or experiences. However, when instructional strategies are aligned with personal experience (specifically as they relate to the use of technology) the focus of a task shifts from the content to the use of technology. Consequently, students become excited about the opportunity to test their skills in familiar ways and view a task as engaging (Heafner, 2004).

Competence

The third aspect of Keller's (1987) theory focuses on students' confidence in their ability to perform a given task. Students must be relatively confident they can be successful when attempting a task to be motivated to make an attempt. If students perceive the task as too difficult, they will avoid it all together (Heafner, 2004). In traditional methods of instruction, students are challenged not only by the content itself but also with asking or answering questions in front of their peers, fearing the possibility of public humiliation (Caldwell, 2007).

Because students are familiar with technology and have a positive feeling about its use, they are likely to have the confidence to address questions asked by the teacher when technology facilitates the exchange (Warschauer, 1996). Further, because technology enables relative anonymity, allows for immediate feedback and provides opportunities for personal control, it motivates students in ways that traditional methods of instruction do not (Fox, 1988; Heafner, 2004; Jacobson & Xu, 2002).

Satisfaction

The last motivating factor in Keller's (1987) model is satisfaction. Satisfaction refers to the degree to which students view their learning experiences as positive. The more positive the experience for the student, the more motivation they have to engage in the learning. When educators use technology-enhanced instruction, as opposed to traditional methods of instruction, research indicates students are more likely to feel part of a community, are more comfortable developing thoughts and ideas, and feel as if they learn from each other (Warschauer, 1996). Additionally, students report feeling a greater sense of personal power, feel less isolated and feel as if technology creates a situation in which it is less threatening to contact people or express their ideas (Heafner, 2004; Warschauer, 1996).

The motivating aspects of learning through technology are widely supported by a vast and growing body of literature (Fox, 1988; Guthrie & Carlin, 2004; Heafner, 2004; Judson & Sawada, 2002; PIALP, 2008; Warschauer, 1996;). Additionally, technology creates opportunities for a more interactive classroom environment (Duncan, 2005; Hake, 1998, 2002; Marrs & Novak, 2004; Meltzer & Manivannan, 2002; PIALP, 2008; Poulis, Massen, Roben, & Gilbert, 1998). Consequently, the explosion of technology's use in learning environments should come as no surprise.

Learning With Technology

The Center for Public Broadcasting (2002) documented students' use of technology for research, tutoring help, and word processing just to name a few. Although technology is motivating to students (Warschauer,1996), it is important to remember that technology alone will not increase student learning. However, when used properly technology can make more effective pedagogy possible and can add value by addressing specific instructional deficits (Beatty, 2004; Draper, 1998; Marrs & Novak, 2004; Duncan, 2005; Meltzer & Manivannan, 2002; Poulis et al., 1998; Suchman, Uchiyama, Smith, & Bender, 2006). Specifically, technology can make it possible to actively engage all students in the learning process and provide immediate feedback to students about their learning.

One piece of technology showing promise is the electronic response system (Judson & Sawada, 2002; Marrs & Novak, 2004; Poulis et al., 1998). Over the past 15 years, the use of electronic response systems have become prevalent on postsecondary campuses and is beginning to penetrate K-12 classrooms (Caldwell, 2007; Fies & Marshall, 2006). While students and teachers alike report positive experiences when working with electronic response systems (Judson & Sawada, 2002; Suchman et al., 2006), the data regarding increases in student learning are mixed (Bunce, Van Den Plas, & Havanki, 2006). Early research reported no academic benefit from such technology (Littauer, 1972). Given the novelty, relative expense at the time and the perceived lack of impact, electronic response systems did not catch on for some time. However, recent studies conducted in postsecondary settings (Caldwell, 2007; Crouch & Mazur, 2001; Hake, 1998; Judson & Sawada, 2002;) found that when used with effective pedagogical strategies, electronic response systems were linked with increases in student learning, although perhaps not more so than when other effective questioning strategies are used with students in the classroom. While the research literature (Caldwell, 2007; Carini et al., 2006; Judson & Sawada, 2002) pointed to multiple factors driving this increase in learning, there is convergence across multiple studies regarding a few of the factors involved. Of particular interest, electronic response systems represent an efficient and low-cost method for increasing engagement through active learning strategies and the use of effective technology appropriate for today's students (Caldwell, 2007; Duncan, 2005; Marrs & Novak, 2004; Novak et al., 1999; Poulis et al., 1998; Smith et al., 2005). Unfortunately, what little literature existed on these systems in a K-12 setting relied heavily on survey methodologies and was largely focused on student/teacher perception or described implementation practices while largely remaining silent as to the related impacts on student achievement (Penuel, Roschelle, & DeBarger, 2006). Because of this deficit, it is necessary to broaden the scope of research models used to explore electronic response technology and its impact on student learning. Consequently, exploring the efficacy of electronic response systems in ninth-grade science classrooms is worthy of empirical study.

In this study, I will explore the impact of electronic response devices on student achievement and student engagement in a ninth-grade science classroom. In exploring these issues, I will address three key questions:

1. Are differences in achievement statistically significant between classes where students use electronic response devices and classes where they are not used?

2. Are differences in overall active engagement statistically significant between classes where students use electronic response devices and classes where they are not used?

3. Are differences in active engagement, specifically during question-and-answer sequences, statistically significant between classes where students use electronic response devices and classes where they are not used?

CHAPTER II

LITERATURE REVIEW

This literature review introduces several constructs and develops linkages between each. First, I identify engagement as a necessary precursor to learning and then define engagement using research literature. Subsequently, the review points to active learning as a way to increase student engagement. Finally, I present electronic response systems as an effective and efficient way to facilitate active learning with varying results regarding their ability to increase student achievement.

Engagement

At all levels of education, considerable effort has gone into determining factors that lead to higher levels of academic achievement. Although broadly defined, research finds engagement positively linked to desirable learning outcomes (Carini et al., 2006; Eckert, 1989; Hake, 1998). The National Research Council's Committee on Increasing High School Students' Engagement and Motivation to Learn (National Research Council and the Institute of Medicine, 2004) released a report on the nature and conditions of engagement, presenting engagement as comprised of both behaviors (e.g., persistence, effort, attention) and emotions (e.g., enthusiasm, pride in success). In terms of behaviors, cognitive behaviors (attention, problem solving, use of metacognitive strategies) were more important than more observable behaviors (completing work, taking more difficult classes, asking for help), as only genuine cognitive engagement resulted in learning.

Newmann's (1992) definition of engagement, as stated in *Student Engagement and Achievement in American Secondary Schools*, has become part of the theoretical empirical framework of this dissertation. Engagement is the student's "psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills or crafts that academic work is intended to promote" (p. 12). Newmann delved beyond the initial layer of motivation—i.e., the desire to do well in school as a means of receiving symbols of high performance such as high grades or approval—to address the inner quality of true concentration and effort required for authentic learning. In an effort to differentiate motivation from engagement, Newmann described how students completed work or performed well in school without truly being engaged in the mastery of a skill, topic or craft—a phenomenon identified in a significant body of research (Eckert, 1989; McNeil, 1986; Weis, 1990). In addition to providing a definition of engagement, Newmann suggested that student engagement results from three factors: (a) students' underlying need for competence, (b) students' perception of membership in the school as a whole, and (c) the degree to which work students are asked to do is authentic.

While Newmann's (1992) work provided a definition of engagement as a construct, Kuh (2001) developed instruments used to measure levels of engagement.

Kuh developed the *National Survey of Student Engagement* (NSSE), which measured the degree to which students devoted time and energy to educational activities known to lead to high levels of student engagement (Chickering & Reisser, 1993; Pascarella &Terenzini, 1991). The NSSE, based on the premise that what students *do* has more of an impact on desired learning outcomes than who they are or where they go to school, uses certain observable behaviors (preparing two or more drafts of a paper, coming to class having completed readings and assignments, working with other students on a project) as a proxy for engagement. The survey asks respondents to indicate how often they engage in those practices. Although initially designed as a tool used to assess the overall quality of an educational program, the NSSE produces data that researchers use to explore connections between measures of student engagement and student learning.

In one such study, Carini et al. (2006) analyzed the relationship between student engagement and learning as measured by (a) RAND tests, (b) essay prompts on the GRE, and (c) college GPA scores. Findings indicated higher scores on the NSSE correlate with higher academic performance as measured by those assessments. The results of this study support other findings (Caldwell, 2007; Judson & Sawada, 2002; Marrs & Novak, 2004; Newmann, 1992) that higher engagement in the learning process yields greater learning outcomes.

Based on research findings, the task of increasing student achievement becomes, first, a matter of increasing student engagement. Newmann (1992) aptly summarized the focus on engagement when he stated that "until we learn more about the fundamental problem of how to engage students in schoolwork, there is no reason to expect improvements in achievement, however those outcomes may be defined" (pp. 3-4). Logically speaking, if greater achievement depends, at least in part, on increased engagement, then the question becomes how to maximize student engagement.

Active Learning

There are certain practices known to lead to higher levels of student engagement (Duncan, 2005; Freeman et al., 2007; Marrs & Novak, 2004; Smith et al., 2005). One of the best-known sets of engagement indicators was presented by Chickering and Gamson (1987). The seven principles include student-faculty contact, cooperation among students, encouragement of active learning, prompting feedback, time on task, high expectations, and respect for diverse talents and ways of learning. It is widely accepted that students experience greater academic achievement when they are actively involved in the learning process (Hake, 1998, 2002; Meltzer & Manivannan, 2002; Smith et al., 2005; Udovic et al., 2002). Active learning strategies rely on the premise that students who actively engage with the material are more likely to recall information (Bonwell & Eison, 1991; Bruner, 1961; Watkins & Slocum, 2007) and that real cognitive engagement is more important than behavioral engagement (Mayer, 2004).

Guthrie and Carlin (2004) suggested today's students are primarily active learners and traditional methods of instruction are therefore increasingly out of touch with student learning styles. An example of research findings supporting this claim is

Hake's (1998) study. Hake analyzed pre/posttest data from 62 introductory Physics classes enrolling over 6,500 students. When comparing averaged normalized gains between classes employing technology-enhanced interactive-engagement methods (IE) and those employing more traditional methods (incorporating little or no use of interactive-engagement methods and relying primarily on passive-student lectures, recipe labs or algorithms), Hake found learning gains in IE classes to be nearly two standard deviations above those in traditional courses.

Similar findings include Mazur's (1997) work using Peer Instruction. Peer Instruction called for students to consider a question, record a response and then defend their answer to peers in small-group discussions. This more active, instructional strategy demanded that students engage the content covered in class and use critical thinking skills as they defended their answers and evaluated their peers' perspectives. Crouch and Mazur (2001) compiled 10 years of results indicating steady increases in student learning, with several studies supporting that claim (Duncan, 2005; Knight & Wood, 2005; Nichol & Boyle, 2003; Rao & DiCarlo, 2000; Smith et al., 2005).

In this study, I theorize that the use of an electronic response device (ERD) constitutes an active teaching method and as such will result in an increase in both student engagement and achievement. In addition to profiting from the active/interactive methodology, students will be more likely to cognitively engage with the content because of their behavioral engagement with familiar technology.

Electronic Response Devices

Understanding the literature focusing on electronic response devices requires attention to the evolution of early findings through modern systems and their use in contemporary instructional settings. Recent research provided valuable insight into the benefits of using electronic response systems with specific, yet mixed, evidence that their use results in learning gains.

Early Findings

Electronic response systems, although not new technology, have only recently emerged as common in classrooms and lecture halls. The first references to electronic response systems are in research literature from the early 1970s (e.g., Bessler & Nisbet, 1971; Casanova, 1971; Chu, 1972; Garg, 1975; Littauer, 1972). Early research focused primarily on the impact that electronic response devices had on student learning. Findings from these studies did not support the claim that response systems increase student learning (Brown, 1972; Casanova, 1971; & Littauer, 1972) and instead presented evidence that no statistically significant difference existed between groups who used response systems and those who did not use them. Despite a lack of evidence pointing to higher learning gains, students in classes where the systems were present indicated a nearly universal positive experience (Bapst, 1971; Brown, 1972; Casanova, 1971; Garg, 1975; Littauer, 1972).

These studies reported students had a positive attitude about the class, found the system useful, had feelings of greater understanding of the content and were more likely to attend class. Given that (a) students indicated a positive experience when using the technology, (b) student performance did not diminished due to its use, and (c) the technology became more sophisticated and less expensive when commercially available from a number of companies, the use of electronic response systems increased over time.

Modern Systems

The first popular, commercially available, response system was *Classtalk*. *Classtalk* was on the market from 1992 until 1999 when it was replaced by simpler, easier to use, easier to support and less expensive response systems like *EduCue PRS* and *eInstruction's CPS* (Beatty, 2004). Today, electronic response devices appear in research literature under several names—e.g., audience response systems, voting machines, wireless keypad response systems, classroom communication systems, immediate response systems and personal response devices.

Despite the advantages and disadvantages unique to particular systems, recent comparisons of commercially available systems (Barber & Njus, 2007; Burnstein & Lederman, 2003) and reviews of current research on response systems (Beatty, 2004; Caldwell, 2007; Fies & Marshall, 2006; Judson & Sawada, 2002; Roschelle et al., 2004) indicate similarities among all systems, each with the same essential attributes.

Condensed to its simplest form, an electronic response system is technology that (a) allows a teacher to project a question, problem or prompt to the entire class; (b) allows students to individually (or in groups) enter their answer using some type of electronic device (referred to throughout as a clicker); and (c) instantly collects and summarizes student responses for both teachers and the students.

The typical response system consists of a handheld transmitter, which students use to submit their answers, a receiver, which collects the responses, a computer that runs the software enabling the collection and summarization of responses, and a projector used to present questions as well as histograms of student responses. Modern transmitters are approximately the size of a remote control with a 10-digit numeric keypad and buttons—e.g., a power button, send button or buttons enabling text entry (Barber & Njus, 2007). Unlike earlier clickers, which were hard wired to the rest of the system, modern transmitters are wireless and are capable of two-way transmission using infrared or radio frequency transmissions. Students submit their responses and get some indication of a received submission. Projectors then display aggregated student responses on a screen for further discussion. While this collection and orchestration of technology may seem complex, it is quite easy to use and requires only moderate technology skills.

Use of Response Systems

Exactly how an electronic response system is used is completely up to the teacher. Reviews of research on the use of response systems indicate their use ranges from a peripheral to a more central role during instructional time (Caldwell, 2007; Fies & Marshall, 2006; Roschelle et. al., 2004). A peripheral role may be inserting the occasional question into an otherwise traditional lecture, using the clicker for facilitating quizzes or using the system to track attendance (Beatty, 2004). Used in this way, electronic response systems do not have an appreciable impact on either student learning or the teacher's understanding of their students' mastery of a given concept. However, when an electronic response system occupies a major role in the instructional strategy, research literature documents multiple benefits (Caldwell, 2007; Crouch & Mazur, 2001; Reay et al., 2008; Roschelle et al., 2004). In such cases, teachers introduce students to new content before class through readings or Internet-based materials with in-class time used to explore student understandings of the content using the clickers. The primary avenue for exploring students' understanding of the content is by asking questions that students answer using clickers. In some cases, students are given the opportunity to discuss their ideas with peers and then answer as a group, but the predominate method for answering questions is the individual response (Beatty, 2004; Burnstein & Lederman, 2003; Fies & Marshall, 2006; Penuel, Roschelle, Crawford, Shechtman, & Abrahamson, 2004). Once received, systems display aggregated answers most often using a histogram without the correct answer being identified. Instructors may then allow students to discuss the distribution of answers, evaluate other students' thinking and/or defend their own ideas. This process can lead to a far more lively exchange than would a more traditional lecture and "raise-your-hand-to-answer" type of pedagogy. It is important to note that positive results in student learning documented throughout research literature are not due simply to the presence of technology (Caldwell, 2007; Judson & Sawada, 2002). According to Beatty (2004), the use of technology does not inherently improve learning, but when used properly it makes effective pedagogy more possible.

Benefits of Clickers

Although research regarding the precise mechanisms driving observed learning increases is mixed (Fies & Marshall, 2006; Roschelle et al., 2004), several factors seem to have a direct impact. These factors are (a) increased student engagement, participation and active learning (Suchman et al., 2006; Duncan, 2005); (b) increased opportunities for providing feedback to students on formative assessments (Black & Wiliam, 1998a, 1998b); and (c) increased feedback to the teacher regarding student understanding (Poulis et al., 1998).

Because technology is motivating for today's students (CPD, 2002) and this type of technology engages students in the content of the class, the use of clickers makes students active in the learning process. As such, clickers possess the potential to increase student learning (Dufresne, Gerace, Leonard, Mestre, & Wenk, 1996; Mazur, 1997). Clickers increase student participation by encouraging all students to respond to all questions asked by the teacher. This technology enables students to offer their responses to a question anonymously and receive anonymous feedback. Hence, they are more likely to attempt answering even challenging questions (Caldwell, 2007; Wood, 2004) without fear of public humiliation by volunteering an incorrect answer. When students take the time to attempt an answer—even if only guessing—they are more psychologically invested in the question and more likely to pay attention to the discussion that follows (Beatty, 2004; Caldwell, 2007; Wit, 2003). Instructors in the Caldwell (2007) study reported that, when using the clickers, students appear to be more active participants in the class by asking and answering more questions, sleeping less, attending more and engaging in more lively discussion. Other important benefits of clicker use included (a) the sense of community and competency, which is critical for engagement (National Research Council and the Institute of Medicine, 2004); (b) increases in student responses to questions (Bullock et al., 2002); and (c) students seeing that either they are correct in their thinking or that they are not alone in their confusion (Knight & Wood, 2005).

A second factor of using electronic response systems is the ease with which teachers can engage *all* students in formative assessments and then give them feedback on their learning (Roschelle et al., 2004). Although instructors use response systems for summative assessments, such systems have the greatest potential for increasing the effectiveness of instruction when used as a way to formatively assess student learning

(Crouch & Mazur, 2001). Formative assessments have been widely studied and referred to as possibly the most effective instructional practice (Bell & Cowie, 2001; Black & Wiliam, 1998a, 1998b).

Formative assessments are defined as assessments used to "enhance rather than evaluate learning" (Beatty, Gerace, Leonard, & Dufresne, 2006, p. 33) and serve to give students feedback about their own learning. Specifically, formative assessments can help students identify limitations to their conceptual understanding and can give students the information needed to become active and efficient learners in the classroom. Admittedly, formative assessments with informative feedback can occur in the absence of any response technology. However, the benefit of an electronic response system is that frequent formative assessments are more efficient and given in such a way that *all* students participate, with immediate feedback available in ways not possible with other low-tech methods. For example, teachers could collect responses by asking students to raise their hands, clap, or use some type of prefabricated manipulation. While all possible, each of these low-tech alternatives lacks the autonomy of a clicker response (potentially impacting the honesty of, and willingness to provide, answers), requires more time and effort in collecting responses (potentially disengaging students) and results in trends in response data observable only to the instructor. The technology addresses the shortcomings of other low-tech, more traditional alternatives and enables frequent questioning and feedback critical to producing higher gains in learning (Black & Wiliam, 1998a, 1998b), even when students initially provide incorrect answers (Guthrie, 1971).

The third beneficial factor of electronic response system use is the immediate feedback available for the teacher. While formative assessments have a documented and direct benefit to students in evaluating their own learning, formative assessments are equally beneficial to teachers in evaluating both student learning and the effectiveness of the instructional approach (Beatty, 2004; Beatty et al., 2006; Caldwell, 2007). Used at the beginning of a unit or lesson, data collected by a response system can help a teacher design an instructional strategy appropriate for the level of understanding and specific misconceptions students bring with them to class. When used consistently, response data provide feedback about students' ongoing learning and emerging conceptions and misconceptions. Effectively employed, this feedback assists the teacher in making adjustments in instruction. Along this vein, several options are available to teachers using response devices that are unavailable when using other low-tech, more traditional methods of formative assessment. For example, teachers can (a) dynamically adjust the focus of a lecture based on trends in student responses, (b) quickly address emerging misconceptions, (c) avoid spending too much time on a particular concept on which students demonstrate solid understanding, or (d) slow down during a lecture as students indicate they are confused. Therein lies the worth of the instructional technology, according to Draper (1998), as it addresses a specific instructional deficit. Although research suggests gains in student learning are linked to effective instructional practices (Bell & Cowie, 2001; Black & Wiliam, 1998a, 1998b), electronic response devices make those instructional practices possible (Beatty, 2004; Crouch & Mazur, 2001; Dufresne et al., 1996; Guthrie & Carlin, 2004; Hake, 1998; Preszler et al., 2007).

Learning Gains

Multiple benefits may be available when implementing electronic response systems as outlined in the previous section. Feedback about the benefits of clickers from teachers and students is quite positive and well documented in the literature (Caldwell, 2007; Draper & Brown, 2004; Judson & Sawada, 2002; Preszler et al., 2007). However, findings vary as to the efficacy of electronic response devices at increasing student learning (Bunce et al., 2006, Crossgrove & Curran, 2008; Suchman et al., 2006).

Bransford et al. (1999) proposed that, theoretically, classroom response technology is one of the most promising advances in educational technology with the greatest capacity to transform classrooms into learner-centered, knowledge-centered, assessment-centered and community-centered places of learning. In a recent study, Roschelle et al. (2004) conducted a meta-analysis of 26 classroom studies employing electronic response systems and found patterns of agreement on the benefits of using such systems. For example, 16 of the studies found evidence of greater student engagement, with 11 of those studies indicating increased understanding of the subject matter. Similarly, Hake (1998) released findings from a survey of 6,542 students' pre/posttest data in an introductory physics course. Findings from the report showed students in courses with electronic response devices, used in concert with interactive teaching strategies, had average learning gains nearly two standard deviations greater than those students in classes where the devices were not used and instead a more traditional instructional approach was employed. Preszler et al. (2007) conducted a study on the effects of student response systems on student learning and attitudes over a range of New Mexico State biology courses. Findings indicated students had positive experiences when using the system and that as the use of the system increased, student learning increased as well. Although not the primary question in the study conducted by Bullock et al. (2002), the average score on exams in an introductory physics course increased from 45% to 75% following the use of an electronic response device. Reay et al. (2008) found that Ohio State University physics students using electronic response devices had significant achievement gains in conceptual learning when compared to students not using the systems. Additionally, this study suggested the use of the system decreased the gap between male and female students' learning gains. Crouch and Mazur (2001) presented 10 years of continuous improvement in pre/posttest gains from students in courses taught using his Peer Instruction (PI) pedagogy, which included the use of electronic response devices. Although other, low-tech methods allowed for formative assessment, electronic response systems increased efficiency and immediacy of the feedback to students and teachers.

While there exists a body of research supporting the role of clickers at increasing learning gains, there is also a body of research suggesting there is little if any impact on learning. Early research found no significant impact when using electronic response systems. Brown (1972) conducted a study with college-level mathematics students using electronic response devices. In that study, the treatment group had instruction bolstered by question-and-answer sessions facilitated through electronic response devices while the comparison group used no such systems. He found no statistically significant difference between the two groups and concluded that students learned as well with electronic response systems as without them. Other, similar studies conducted in college-level chemistry, physics and economics courses yielded the same statistically insignificant results when testing the impact of electronic response devices (Bessler & Nisbet, 1971; Casanova, 1971).

More recently, Bunce et al. (2006) conducted a study comparing the effect of a student response system on student achievement to its effect on online WebCT quizzes. The study found that student (electronic) response devices had no impact on student learning when measured by performance on teacher-written exams. Suchman et al. (2006) designed a study comparing two groups of students in a microbiology class. Although there were higher test scores in the class that used clickers more extensively, the increase in test scores was the same even when test questions were on material not covered when using the clickers. Paschal (2002) completed a study in a physiology course using an electronic response system and found no statistically significant difference between classes using and those not using the technology. Knight and Wood (2005) found that in many cases the use of an electronic response system had no effect on test scores for students in their Biology courses. Crossgrove and Curran (2008) published an article describing a study in biology and genetics courses using electronic response devices. In that study, they found the use of clickers had no overall effect on exam scores when comparing separate cohorts of students and that there was no statistically significant difference in genetics students' retention of concepts over time when comparing clicker and nonclicker groups of students.

Although findings in recent studies (Bunce et al., 2006; Caldwell, 2007; Suchman et al., 2006) mirror the those of earlier studies suggesting the technology has no impact on student learning, there is a growing number of recently published studies suggesting an electronic response system is associated with multiple benefits for students, including increased student achievement (Bullock et al., 2002; Crouch & Mazur, 2001; Hake, 1998; Reay et al., 2008). Despite an increased research interest in the subject in recent years, reviews of current literature indicate a lack of rigor and limited research methodologies making it impossible to develop strong conclusions about the efficacy of the technology, pointing to a need for more systematic research efforts (Caldwell, 2007; Fies & Marshall, 2006; Penuel et al., 2004).

Gaps in the Research

Electronic response systems are prevalent in classrooms and lecture halls with continued use likely increasing into the future. One indicator of the likelihood for continued use is the fact that the National Science Foundation committed more than $11 million in support of the development and study of this instructional technology (Penuel et al., 2004).

Although predominately found in higher education settings, electronic response technology is now penetrating K-12 classrooms. For example, one commercial vendor, *eInstruction*, claims a distribution of over one million response pads to students in elementary, middle and high schools with more than 1,000 schools listed on their website as clients. Like the response systems themselves, research about electronic response systems is overwhelmingly focused on higher education, with relatively little literature describing the implementation or learning outcomes related to this technology in K-12 settings (Penuel et al., 2004).

Furthermore, regardless of the educational setting, the current body of research is largely descriptive, focuses on student perceptions, relies heavily on surveys and lacks the rigorous systematic approach necessary to guide educational decisions in this area (Caldwell, 2007; Fies & Marshall, 2006; Penuel et al., 2004). Because it is necessary to broaden the scope of research models used to explore electronic response technology and its impact on student learning and because relatively little research has been conducted focusing on outcomes in K-12 settings, exploring the efficacy of electronic response systems in ninth-grade science classrooms is worthy of empirical study.

CHAPTER III

METHODOLOGY

School Setting

The study took place in ninth-grade science classrooms in a comprehensive high school of approximately 1,500 students situated in a town of approximately 50,000 residents in the Pacific Northwest. The demographics of the school are rapidly changing with an influx of students of color, specifically Hispanic students. The student body was comprised of 88.9% Caucasian students, 1.1% African American students, 6.3% Hispanic students, 2.1% Alaskan/Native American students and 1.6% Asian/Pacific Island students. There were 69 teachers at the school, 76.8% having a master's degree or higher, with an average of 14.6 years of experience. The school operated on a trimester schedule, with students taking five 70-minute classes each day.

According to the Oregon Department of Education (2007), the school had a 95% graduation rate, a 1% dropout rate and an attendance rate of 91.3%, with 38% of the students taking the SAT. For the 2006/2007 school year, neither the district nor the school made Adequate Yearly Progress (AYP) as defined by the No Child Left Behind Act of 2001. Results on the Oregon state assessments indicated 57% met or exceeded the reading standards, 43% met or exceeded the math standards and 54% met or exceeded the writing standards.

Description of Participants

The participants in this study included freshmen (ninth-grade) students. Students enroll in this high school from four middle schools within the district. The four feeder middle schools ranged in size from one K-8 school with 166 students to the largest middle school in the district with 637 students. Students from two of the four middle schools go to one of two comprehensive high schools in the district, with their distribution based on neighborhood boundaries. According to the Oregon Department of Education (2007), one of the four middle schools made Adequate Yearly Progress (AYP), while the others did not, with overall rankings ranging from satisfactory to strong on the Statewide Accountability System.

The freshman class consisted of 425 students: 86.6% Caucasian, 1.8% African American, 1.8%Asian/Pacific Islander, 7.3% Hispanic and 2.5% Alaskan/Native American. The sampling frame for this study consisted of the 425 students in the freshman class. From the sampling frame of all freshmen, a convenience sample of all students enrolled in Physical Science (Science A) was used.

Students selected for participation in the study were in freshman Physical Science classes. The Physical Science A course was offered in the morning and served 62 students. All study participants participated in the same district-approved middle school science curriculum: one year of Earth, one year of Life and one year of Physical Science. The science curriculum for the high school in this study required first-year students to take both sections during their freshman year. Each section took place in one of two 12-week terms, with students attending 70-minute classes each school day. The curriculum included lectures, lab activities and assignments designed to help students master the content standards as outlined by the state Department of Education. Content covered in the physical science classes focused on topics such as force and motion, friction, Newton's laws, potential and kinetic energy, power and work, electromagnetic energy, simple and compound machines and gravity.

Research Design

This study used a nonequivalent comparison group pretest/posttest design. Such a methodology allowed for straightforward evaluation of an instructional intervention—in this case, use of electronic response systems—across two points in time in a relatively nonintrusive way without requiring advanced statistical analysis methodologies not normally used by education practitioners. Specific to this study, the teacher administered the Science A pretest to both Science A classes the first day of the second week of fall term. Administration of a posttest took place 11 weeks after the pretest, at the end of the term. The pre- and posttests contained the same items and were administered in identical formats, respectively.

Procedures

During the academic year before data collection, the teacher participating in the study received an Infrared Frequency electronic response system manufactured by *eInstruction*. An Epson projector, used to display questions and aggregated responses, was connected to a Macintosh laptop computer used to run *eInstruction's Classroom Performance System* (CPS) software application. In addition to the software application, the system included a receiver attached to the laptop computer with enough individual handheld electronic response devices for a class of 35 students.

The teacher, along with other members of the Science Department, received training on the use of this electronic response system. The training consisted of two 8-hour work sessions. The first work session focused on system operation, developing familiarity with the software application, entering questions, displaying results and troubleshooting technical problems, and the final 2 hours of the work session were devoted to developing question sequences for content being covered at the time. The second 8-hour work session occurred after teachers had been using the systems for several months. Instructional technology support staff facilitated the work session.

During this session, teachers shared their experiences, successes and difficulties when using the systems and discussed ways to modify instructional practices to make the best use of this technology. Both training sessions occurred before the study began. In addition to discussing technical issues specific to the operation of the electronic response system itself, teachers read and discussed an article on designing effective questions when teaching with electronic response systems (Beatty et al., 2006). Students in the Treatment group learned to use the devices and interpret resulting data during the initial phase of the study.

The study occurred throughout the first trimester, following the first week of the term. Data collection began after the first week of the term in an effort to maximize the potential for an intact cohort by avoiding student mobility due to schedule changes and because instruction during the first week of the term is focused mainly on classroom expectations and procedures largely unrelated to science content learning.

Following the first week of fall term, instruction for the Comparison group included traditional practices such as daily preplanned questions about the content, with students volunteering answers by raising their hands. In addition to homework and lab activities, study participants took planned chapter/unit tests.

Similar to the instructional practices for the Comparison group, participants in the Treatment group had the same assigned homework and lab activities while taking the same chapter/unit tests. Although the teacher asked the same daily preplanned questions of both the Treatment and the Comparison groups, the electronic response system was the mechanism for facilitation of daily question-and-answer practices.

Measures

One measure of interest was the Science A pretest/posttest. Prior to the academic year during which data was collected, a panel of teachers from the Science Department and a doctoral student from the University of Oregon reviewed the test to ensure questions adequately sampled the content domains, were aligned to state science standards, were appropriate for the age and skill level of students in the class and were of an appropriate level of difficulty. Teachers developed the test using a structured response format consisting of 60 questions. Students submitted answers using a Scantron answer sheet, and grading was completed using a Scantron grading machine—thus ensuring objectivity with regard to test scores with no partial credit given. During test administration, all tests were numbered, returned to the teacher and inventoried before dismissing students in an effort to ensure test security.

Reliability of Measure

In the development of any instrument, reliability is the primary statistical procedure for ensuring that inferences based on the data are consistent. Reliability serves as evidence that an instrument is consistently measuring a given variable. However, reliability estimates cannot serve as evidence an instrument is measuring what it is supposed to be measuring. For example, while tallying the number of correct words read per minute can be measured very consistently, such a measurement would not be an appropriate way to measure computational fluency in mathematics. In short, reliability estimates do not ensure a valid inference. Reliability calculates consistency, while validity addresses appropriateness of decisions made from the data collected.

For this data set, internal consistency is the best indicator of reliability. Internal consistency is a measure based on the correlations between different items on the same test. An analysis of internal consistency determines whether several items proposing to measure the same general construct produce similar scores. In this study, the researcher measured internal consistency using Cronbach's alpha (α), a statistic calculated from the correlations between items. Results from the analysis are shown in Table 1. Generally, α increases when the correlations between the items increase. Commonly accepted guidelines are that an α of 0.7 indicates minimally acceptable reliability and 0.8 or higher indicates good reliability. It is critical to note that extremely high reliabilities (0.95 or higher) are not necessarily desirable, as this may indicate item redundancy.

Initial Development of Pretest Measures

Because the freshman science curriculum at the school in this study has two distinct science tracts (physical science and chemical science), the teacher developed two different pretest instruments before the study. The two test instruments, one specific to the physical science content (Science A) and another for the chemical science content (Science B), were administered to the school's freshmen.

Following administration of the two different pretests, I analyzed the internal consistency of both test instruments. Internal consistency of the Science A pretest was adequate (α = .80); this was not true for the Science B pretest (chemical science). Consequently, the classes participating in the Science B curriculum were not included in this study. Although this decision resulted in a smaller sampling population for the study, it was better than the alternative of making inferences based on data collected using an unreliable instrument.

<u>Internal Consistency Data for the Pretest/Posttest</u>

As Table 1 indicates, overall internal coefficients demonstrate good reliability for the groups remaining in the study. Table 1 reflects a range in pretest α from .79 for the Treatment group to .82 for the Comparison group, with an overall value of .80. Reliability estimates for the posttest scores are somewhat higher. Values for α range from .91 for the Treatment group to .88 for the Comparison group, with an overall value of .90.

TABLE 1. Internal Consistency

Group	α
Pretest	
Overall	.80
Comparison	.82
Treatment	.79
Posttest	
Overall	.90
Comparison	.88
Treatment	.91

<u>Assigning Science A Classrooms to Treatments</u>

From the classes in the Science A group, the researcher randomly assigned a class to the Treatment group, with the remaining class serving as a comparison. The

Treatment group progressed through the ninth-grade science curriculum using electronic response technology on a daily basis, while the Comparison group progressed through the same curriculum with more traditional classroom practices such as raising hands to volunteer answers to teacher questions.

Measures of Engagement

Another critical measure was student engagement. Using on-task behavior and attendance rates as proxies, I measured student engagement for students in both groups. To quantify on-task behavior, momentary time sample data were collected through classroom observations in the Treatment and the Comparison groups during Weeks 2, 4, 6 and 8. In each of the classes, data were collected for all students in attendance the day of the observation. A trained observer recorded on-task behavior data at 10-second intervals for one full minute before moving on to the next student, with data collection occurring throughout the entire class period. Student behavior was recorded as falling into one of three categories: (a) on task, (b) passive engagement or (c) off task. I defined on-task behavior, given a value of 2, as characterized by active involvement in a prescribed activity, answering teacher questions or behavior otherwise clearly indicating the student was paying attention during instruction. I defined passive engagement, given a value of 1, as characterized by marginal involvement in a prescribed activity or behavior, while not necessarily disruptive but also not indicative of overt involvement with instruction at the time. I defined off-task behavior, given a value of 0, as characterized by behavior indicating the student was clearly not paying attention to teacher questions or instruction or was clearly not actively involved in the classroom's prescribed activities.

Fidelity of Implementation

Fidelity of implementation was a measure of great importance. Specifically, the frequency of clicker use as well as the consistency of the questions asked to students in both the Comparison and the Treatment group were of interest. To ensure fidelity, the researcher conducted classroom observations of both the Treatment and the Comparison group during Weeks 2, 4, 6 and 8. For classes in the Treatment group, the researcher collected data regarding whether or not the students used clickers during instruction. For classes in the Comparison group, data collection focused on whether or not the teacher asked the same preplanned questions of students in both the Treatment and Comparison groups.

Data Analysis

Findings regarding knowledge acquisition resulted from statistical analysis of the test scores. Differences in the pre and posttest scores were analyzed using a repeated-measures ANOVA with between-subject factor. Admittedly, an issue to address when conducting the ANOVA was the unbalanced group sizes, which violates an underlying assumption of the ANOVA statistic. While I acknowledge that unequal sample sizes can produce confounds, my sample size difference of two students did not violate North Carolina State University Professor Garson's (2009) *rule of thumb*. On his website, Garson wrote that an

> ANOVA is robust for small and even moderate departures from homogeneity of variance (Box, 1954, 1978). Still, a rule of thumb is that the ratio of largest to smallest group variances should be 3:1 or less. . . . Marked violations of the homogeneity of variances assumption can lead to either over- or under-estimation of the significance level.

Garson's claim is substantiated by Halpin, Carwile, and Halpin (1991). They declared that in nonexperimental research, unequal cell sizes usually reflect reality. They contended that (a) "unbalanced factorial designs are the most widely employed ANOVA design;" (b) "it is the rule rather than the exception for researchers to have unequal cell sizes in their investigations;" and (c) "[they] tend to be skeptical of the investigations where researchers report equal cell sizes, especially when they fail to explain procedures utilized to obtain equal cell sizes" (p. 1).

As previously mentioned, momentary-time-sample (on-task behavior) data and attendance rates were used as proxies for student engagement. The study employed an independent samples *t*-test to analyze the differences between groups' on-task behavior both throughout the entire class period as well as specifically during question-and-answer portions of instruction. Finally, an independent samples *t* test was used to determine if a statistically significant difference existed in attendance rates between the Treatment and Comparison groups.

CHAPTER IV

RESULTS

Following a presentation of the demographic statistics and average attendance data, I present the results of Cronbach's (α) test for internal consistency. Next, I present the descriptive statistics for the pre/posttests as well as the analysis of variance results for student scores on the test instrument used in the study. A graphical representation of growth for both the Treatment and the Comparison groups accompanies the associated tables of data. Subsequently, tables of data regarding the differences between groups on levels of engagement across a class period accompany a visual display of the average mean scores for student-level on-task behavior. Finally, tables of data regarding the differences between groups on levels of engagement across teacher questioning accompany a visual display of the average mean scores for student-level on-task behavior across teacher questioning.

Student Demographic Data

Table 2 presents the demographic statistics of the study participants. The Comparison group consisted of 31 students; 18 were male and 13 were female. The average number of days missed was 2.46. The Treatment group consisted of 29 students; 17 were male and 12 were female. The average number of days missed was

3.65. An independent samples *t* test indicated that no statistically significant difference ($p = .24$) existed between the two groups' attendance rates. The average age for students in the Comparison group was 13.67, whereas the average age for students in the Treatment group was 13.59. An independent samples *t* test indicated no statistically significant difference ($p = .51$) in average student age between the groups.

TABLE 2. Demographic Statistics

Group	*N*	Average age	Gender	Average days missed
Comparison	31	13.67	Males = 18 Females = 13	2.46
Treatment	29	13.59	Males = 17 Females = 12	3.65

Differences Between Groups After Intervention

The first question asked whether there would be a significant difference between the Treatment and Comparison groups on their mean academic scores because of the Treatment group's use of the electronic response device.

Table 3 presents the descriptive statistics for the pre- and posttest scores. The Comparison group ($n = 31$) had a pretest mean score of 31.65 with a standard deviation of 7.37. The Treatment group ($n = 29$) had a pretest mean score of 29.93 with a standard deviation of 7.18. For the posttest the Comparison group's mean score increased to a value of 49.29 with a standard deviation of 7.61, while the Treatment group's mean

TABLE 3. Descriptive Statistics

	Group	*M*	*SD*
Pretest	Comparison (*n* = 31)	31.65	7.37
	Treatment (*n* = 29)	29.93	7.18
Posttest	Comparison	49.29	7.61
	Treatment	50.97	8.01

score increased to a value of 50.97 with a standard deviation of 8.01. Table 3 contains the complete descriptive statistics.

Figure 1 below graphically displays the changes in mean test scores over time. The line chart shows that each group experienced similar growth over time.

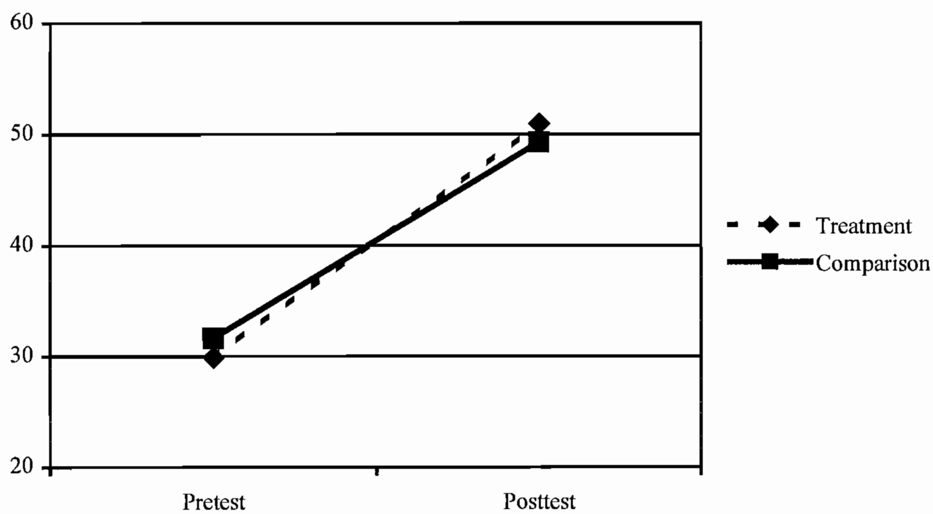

FIGURE 1. Change in mean score over time.

Table 4 presents the results from the repeated measures ANOVA conducted for test scores. A main effect for time (pretest to posttest) was found, $p < .001$, indicating a statistically significant growth rate, with a medium effect size, $d = .40$. However, there was no main effect for group, $p = .99$. Thus, no interaction between group (Treatment versus Comparison) and time (pretest versus posttest) was found, $p = .13$. See Table 4 for complete ANOVA statistics.

TABLE 4. Analysis of Variance Results

Source	Type III Sum of Squares	*df*	Mean Square	*F*	*p*-value	Partial Eta Squared
			Tests of within-subjects contrasts			
Time	11208.39	1	11208.39	303.35	.00	.84
Time x Group	86.06	1	86.06	2.33	.13	.04
Error (time)	2143.03	58	36.95			
			Tests of between-subjects effects			
Intercept	196203.81	1	196203.81	2549.65	.00	.98
Group	.01	1	.01	.00	.99	.00
Error	4463.28	58	76.95			

Differences Between Groups on Levels of Active Engagement Across the Class Period

The second question asked whether there would be a significant difference between the Treatment and Comparison groups on their mean levels of active engagement because of the Treatment group's use of the electronic response device.

Table 5 presents mean scores for student-level on-task behavior across an entire class period in both the Treatment and Comparison groups. Those data helped to determine the average on-task behavior value at each observation interval for students in both the Treatment and Comparison groups. With respect to on-task behavior values at each observation interval, the Treatment group averaged 1.83 ($SD = 0.01$), while the Comparison group averaged 1.70 ($SD = 0.04$). See Table 5 for complete Observation statistics.

TABLE 5. Observation Means

Group	*n*	*M*	*SD*	Std. Error Mean
Comparison	31	93.42	13.27	2.38
Treatment	29	107.17	7.62	1.42

Figure 2 visually displays the average mean scores for student-level on-task behavior for both the Treatment and Comparison groups across an entire class period. Visually, one can see that the mean Treatment group score was higher than the Comparison group's means score.

Table 6 presents results for the independent samples *t* test comparing differences in on-task behavior throughout an entire class period between the Treatment and Comparison groups. The findings show statistical significance, $p < .01$. There was a very large effect size, $d = 1.28$, indicating the educational significance of these

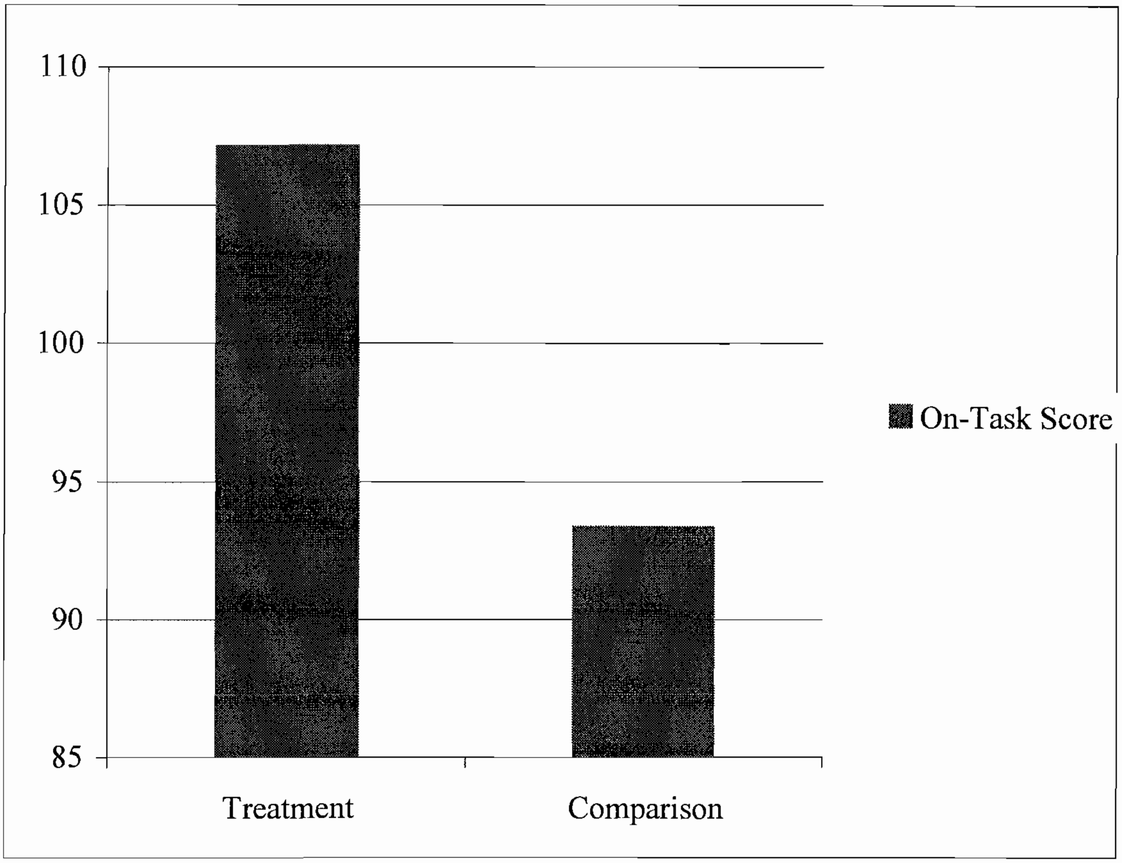

FIGURE 2. On-task behavior values at each interval.

TABLE 6. Independent Samples *t* Test: Class Period

	t	*df*	Sig. (2-tailed)	Mean difference	Std. Error difference	95% Confidence Interval of the Difference	
						Lower	Upper
Equal variances not assumed	-4.96	48.44	.00	-13.75	2.77	-19.32	-8.18

differences. With an effect size of this magnitude the average student in the Treatment group would be ranked in the 96.4th percentile of the Comparison group with respect to on-task behavior (77.4% nonoverlap). See Table 6 for complete statistics.

Differences Between Groups on Levels of Engagement Across Teacher Questioning

The third question asked whether there would be a significant difference between the Treatment and Comparison groups on their mean levels of active engagement during teacher questioning because of the Treatment group's use of the electronic response device.

Table 7 presents mean scores for student-level on-task behavior specific to observations during which the teacher was asking daily preplanned questions. Those data helped to determine the average on-task behavior value at each observation interval. With respect to on-task behavior values at each observation interval during question-and-answer sessions, the Treatment group averaged 1.97 ($SD = 0.006$), while the Comparison group averaged 1.49 ($SD = 0.39$). See Table 7 for complete statistics.

TABLE 7. Observation Means

Group	*n*	*M*	*SD*	Std. Error Mean
Comparison	17	13.29	6.56	1.59
Treatment	22	21.64	8.89	1.90

Figure 3 visually displays the average mean scores for student-level on-task behavior for both the Treatment and the Comparison groups during which the teacher was asking daily preplanned questions. Visually, one can see that the mean Treatment group score was higher than the Comparison group's mean score.

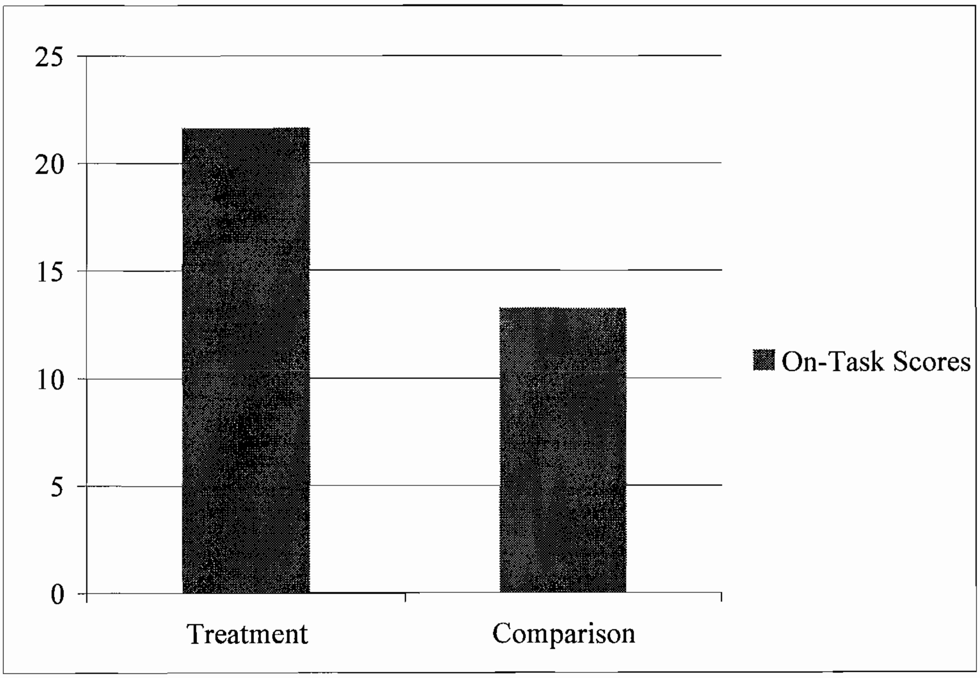

FIGURE 3. On-task behavior values at each observation interval during questioning.

Table 8 presents findings from the independent samples *t* test comparing differences in mean scores for on-task behavior specific to observations during question-and-answer sessions of the class. Results indicate a statistical significance, $p =$ .003. There was a large effect size, $d = 1.08$, indicating the educational significance of differences in scores. With an effect size of this magnitude the average student in the

Treatment group would be ranked in the 86th percentile of the Comparison group with respect to on-task behavior (58.9% nonoverlap). See Table 8 for complete statistics.

TABLE 8. Independent Samples *t* Test: Teacher Questioning

	t	*df*	Sig. (2-tailed)	Mean difference	Std. Error difference	95% Confidence Interval of the Difference	
						Lower	Upper
Equal variances not assumed	-3.37	36.94	.00	-8.34	2.47	-13.36	-3.33

Results Summary

A test of Cronbach's alpha (α) determined the test instrument's internal consistency to be adequate for the Science A course. The values ranged from a pretest α of .79 for the Treatment group to .82 for the Comparison group with an overall value of .80. Reliability estimates for the posttest scores were somewhat higher. Values for posttest α ranged from .91 for the Treatment group to .88 for the Comparison group with an overall α value of .88 (see Table 1).

The descriptive statistics for the pre- and posttest scores indicate growth for both groups over the course of the study. The Comparison group, $n = 31$, had a mean pretest score of 31.65 ($SD = 7.37$) that increased to a mean posttest score of 49.29 ($SD = 7.61$), whereas the Treatment group, $n = 29$, had a slightly lower mean pretest score of 29.93 ($SD = 7.18$) that increased to a mean posttest score of 50.97 ($SD = 8.01$). The repeated measures ANOVA used to analyze posttest scores yielded a significant difference for time, $p < .01$, no main effect for group, $p = .99$, and thus no significant interaction for time-by-group, $p = .13$.

On-task behavior scores reflect the mean score for the class, the mean score for the observation interval, and the *t* tests results comparing the differences between the Treatment and Comparison groups. Data were reported for the overall class period during which the observation took place as well as within the time specifically devoted to teacher question-and-answer practices.

The overall mean score for the Comparison group was 93.42 ($SD = 13.27$), whereas the overall mean score for the Treatment group was 107.17 ($SD = 7.62$). With respect to on-task behavior values at each observation interval, the Treatment group averaged 1.83 ($SD = 0.01$), while the Comparison group averaged 1.70 ($SD = 0.04$). The independent samples *t* test comparing differences in on-task behavior throughout an entire class period between the Treatment and Comparison groups resulted in a significant difference, $p < .01$. This significant difference also yielded a very large effect size, $d = 1.28$.

The mean score for on-task behavior in observations during which the teacher was asking daily preplanned questions was 13.29 for the Comparison group ($SD = 6.56$) and 21.64 for the Treatment group ($SD = 8.89$). With respect to on-task behavior values at each observation interval during question-and-answer sessions, the Treatment group averaged 1.97 ($SD = 0.006$), while the Comparison group averaged 1.49 ($SD = 0.39$). The independent samples *t* test comparing differences in mean scores for on-task behavior specific to observations during question-and-answer sessions of the class resulted in a significant difference, $p = .003$, and a large effect size, $d = 1.08$.

CHAPTER V

DISCUSSION

I designed this study to explore the role electronic response systems play in increasing student learning in ninth-grade science classrooms. In exploring the issue, I addressed three key questions:

1. Are differences in achievement statistically significant between classes where students use electronic response devices and classes where they were not used?

2. Are differences in overall active engagement statistically significant between classes where students use electronic response devices and classes where they were not used?

3. Are differences in active engagement specifically during question-and-answer sequences statistically significant between classes where students use electronic response devices and classes where they were not used?

Review of Findings

Student demographic data indicate comparability between the groups. While the pretest-posttest findings reflect significant growth over time for students in both the Treatment and the Comparison groups, ANOVA results showed a main effect for time, $p < .01$, no main effect for group, $p = .99$, and thus no interaction effect, $p = .13$. Active engagement data revealed statistically significant differences between the groups across an entire class period and more specifically across teacher questioning. The Treatment group demonstrated significantly higher levels of engagement across the class period when compared to the Comparison group. Likewise, levels of engagement across teacher questioning were statistically higher for the Treatment group than for the Comparison group.

Threats to Internal and External Validity

Before this chapter addresses the research questions directly, it is important to review specific components of the design model incorporated as a way of addressing potential threats to internal validity. Of specific concern were (a) group selection; (b) instrumentation; (c) the study's setting—i.e., a single teacher's classroom; (d) fidelity when implementing the treatment; and (e) data types and collection method.

Group Selection

Because the setting for the study was a classroom, I had relatively limited control over the study participants beyond ensuring they were ninth-grade high school students taking a science course. Consequently, random assignment of students to a group was simply not an option. Although the strength of findings from an experimental design are greater than from a quasi-experimental design, I had to work within the scheduling realities of a large high school with a comprehensive curriculum. Therefore, randomly assigning participants to groups was not possible. However, the study's pre-posttest design allowed me to test the assumption of inequality on the outcome variable between groups.

The demographic data in Table 2 reflect a relatively even distribution between the groups in terms of overall numbers of students across groups, gender balance and age of participants within each group. Because all participants in the study came from the same middle school feeder system, they all progressed through the same middle school science program with a district-approved science curriculum. All of the students in the study were in their first year of high school and therefore would have felt the impact of the transition to high school in relatively similar ways. This is an important point given the research literature regarding the academic and social ramifications of the transition to high school (Hayes, Nelson, Tabin, Pearson, & Worthy, 2002). The two groups of students were equivalent in their academic history, knowledge of science content before the study, and degree of maturation throughout the study.

While it may be possible for readers to generalize within this study, any generalization to other settings should be done cautiously. My findings can generalize only to students from similar age groups studying similar content and enrolled in like high schools. Because many of the challenges facing first-year high school students are unique to that transitional period, it cannot be said, with any degree of certainty, that, if replicated for senior-level students, the study would yield similar results. Although similar results might be possible, one would still need to design a rigorous study to generate valid data.

Likewise, the school size and school setting need to be similar in order to generalize these findings to another setting. The potential confounds introduced by a much larger, or smaller, school would make it difficult to predict exactly how electronic response technology integration would play out. Confounds introduced by a more rural or urban setting, by a school with more or fewer feeder schools or schools with a significantly different socioeconomic composition, would make it difficult to predict the impact of this technology with any confidence.

Instrumentation

Threats to internal validity result from inconsistencies with the test instrument (e.g., different items on a posttest than on a pretest, different scoring rubric from test to test, or a completely different test format, etc.). Given the possibility of such a threat, the instrument design was important. The test (see Appendix) was teacher-developed. Although teachers often develop and administer their own tests without rigorous scrutiny, certain precautions were taken to maximize the validity and reliability of the instrument used in the study. A panel of teachers from the Science Department at the school and a doctoral student from the University of Oregon reviewed the test to ensure that questions (a) adequately sampled the class's content domains, (b) aligned with state science standards, (c) represented an appropriate skill level for the age of the students in the class, and (d) represented appropriate variance in item difficulty. Both the pretest and the posttest were identical structured-response instruments consisting of 60 questions. In each test administration, students submitted answers using a Scantron answer sheet. The teacher graded the tests using a Scantron grading machine, thus ensuring objectivity in test scores with no partial credit given. As a result, the Science A test was adequately vetted and was a valid instrument.

I conducted a test of Cronbach's alpha (α) to determine internal consistency. The values ranged in pretest α from .79 for the Treatment group to .82 for the Comparison group with an overall value of .80. Reliability estimates for the posttest scores were somewhat higher. Values for α ranged from .91 for the Treatment group to .88 for the Comparison group with an overall value of .88. In both the pretest and the posttest, results of the Cronbach's α showed the instrument demonstrated high levels of internal consistency and, as such, should reliably measure the chosen variable (science content knowledge) over time. The design model largely controlled for threats to internal validity due to instrumentation.

Single Teacher

One of the stronger points of the research design is that the study took place in a single teacher's classroom. I designed the study to answer a very specific and germane question for the faculty and administration at this particular school. Namely, how effective are electronic response devices at increasing student achievement in these particular classes? Therefore, a single classroom setting was quite appropriate.

Because the study was limited to a single teacher's classroom, certain threats to validity did not exist. For example, aside from the presence or absence of the electronic response devices, the instruction was essentially the same between the groups. Every teacher, no matter how well aligned with others in the same content area or even teaching the same courses, has his or her own unique approach and instructional routine filled with nuance. All of the homework, lab activities, lectures, grading standards and class-wide discussions were essentially the same between groups. Additionally, teacher-student interaction outside of question-and-answer sessions was the same, as was access to the teacher outside of the regularly scheduled class times. In this model, I eliminated those confounding variables altogether. Consequently, it was easier to associate any difference in student performance or engagement to the treatment. However, a study involving a single teacher at a single site is limited and findings should be interpreted cautiously.

Fidelity of Implementation

Rigorous adherence to the design model was critical to minimize threats to internal validity. I conducted observations of the Treatment and Comparison groups in Weeks 2,4,6 and 8. During each observation period, I collected data on the frequency of clicker use for the Treatment group and the comparability of questions asked in both groups. During every observation session, students in the Treatment group responded to questions using the clickers and the teacher asked the same preplanned questions of both the Treatment and Comparison groups.

The potential consequences of marginal or inconsistent use of the electronic response devices would threaten the validity of any inference made from the findings of this study. In the event of significantly higher posttest scores or measures of engagement, one could incorrectly infer a cause-and-effect relationship between scores and the treatment even though the treatment represented only a marginal role in the instructional routine. The opposite could apply for significantly lower test scores or measures of engagement. Specifically, one could infer the treatment had little or no impact on student achievement when, in fact, clickers were not consistently used in the instructional routine. It was important to observe students consistently using the clickers in question-and-answer sessions on a daily basis.

Threats to validity would exist if the teacher presented a different set of questions in the Treatment classroom than in the Comparison classroom. Research studies (Beatty et. al., 2006; Bullock et. al., 2002; Reay et. al., 2008) point out that the type of questions teachers ask students have as much or more to do with knowledge acquisition as any instructional technology. Were the teacher to ask more diagnostic questions to one group as compared to the other, s/he might get a better sense of emerging misconceptions and better address those student-learning needs. Potential threats to validity could arise if the teacher consistently asked one group sets of questions requiring more critical thinking and the other group sets of questions requiring only lower order thinking. In this scenario, it is possible for any difference in posttest performance to be misinterpreted as due to the clicker treatment when in actuality the difference might be due to the sophistication of the questions asked of each group. In each observation of the Treatment and Comparison groups, the teacher asked the exact same set of preplanned questions. Therefore, the design model controlled such confounding variables.

It is important to notice that the study's design minimized the full compliment of teaching strategies (e.g., rejoinders, peer instruction, forward feedback, etc.). Limiting those factors was necessary and reasoned to ensure that this study determined the degree to which the use, or absence, of the clicker impacted the learning and behavior of the students in each group. While there have been studies involving clickers that also incorporated other interactive teaching strategies (Crouch & Mazur, 2001; Hake, 1998), the reader is again reminded that the purpose of this study was to estimate the impact that clickers *alone* would have on student learning and student engagement without having to parse out other confounding factors.

Data

The types of data collected in this study were sensitive to the theoretical-empirical framework on which the study was built and avoided potential threats to internal or statistical conclusion validity. The study was built on the following theoretical empirical framework: (a) engagement is a necessary precursor to learning, and as such, increases in learning require an increase in student engagement (Caldwell, 2007; Judson & Sawada, 2002; Newmann, 1992); (b) active learning strategies are successful at increasing student engagement (Duncan, 2005; Freeman et al., 2007; Marrs & Novak, 2004; Smith et al., 2005); (c) use of instructional technology (specifically, electronic response devices) promotes active learning, thereby increasing class-wide engagement (Guthrie & Carlin, 2004; Hake, 1998); consequently (d) the use of electronic response devices results in higher learning gains. If I had collected and analyzed student scores only on the test instrument, I would have ignored the treatment's impact on the mechanism driving increased learning—namely, engagement. With this design, I was able to evaluate the impact of the treatment on student engagement as well as student learning gains. Subsequently, I was able to evaluate the connection between these two constructs in this particular setting.

Differences Between Groups After Intervention

In answer to whether or not there was a statistically significant difference in achievement between classes where students use electronic response devices and classes with more traditional instruction, I was unable to show significant differences for clickers when all other instructional variables were held constant.

Although results indicate significant growth over time (pretest to posttest) for both the Treatment and Comparison groups, $p < .001$, there was no interaction effect, p

= .13. While all students in this study demonstrated significant learning gains over time, students using clickers did not achieve at significantly higher levels compared to students who did not use clickers. Findings from this study are consistent with studies conducted in higher education classrooms (Brown, 1972; Bapst, 1971; Crossgrove & Curran, 2008; Paschal, 2002), where clickers had no significant impact on student achievement. A cursory glance at the results suggests no educational benefit from clicker technology. However, a more critical analysis of the results, and the research methodology, explains possible mechanisms driving high achievement in the Comparison group, thus revealing the source of equitable achievement over time; both groups benefited from a foundation of high-quality instruction.

Dissecting this notion first requires an examination of the questions the teacher designed for the students in her classes. As part of the training when first receiving the clickers, the teacher reviewed research (Beatty et. al., 2006) describing the attributes of effective question design when using electronic response technology. The research explored ways questions could target specific aspects of abstract constructs, key in on problem-solving skills, develop conceptual understanding, identify misconceptions and challenge students to engage in critical thinking. The premise of the study was that students' level of content mastery is dependent on the quality and frequency of questions asked them by the teacher. The teacher in this study developed clicker questions adhering to the principles described in that research.

The design model of this study required the teacher to have daily preplanned questions for the Treatment group. To control for question quality and frequency, the teacher had to ask the same number of identically worded preplanned questions to both the Treatment and Comparison groups. Because students in both groups had the same teacher, they benefited from the same pace, activities and student-centered approach adopted by the instructor. Additionally, both groups benefited from a daily battery of preplanned, well-designed questions. The only difference between the classes was the presence, or absence, of the electronic response system. Due to the assertion that technology is simply a tool (Draper, 1998), effective only in conjunction with already effective teaching strategies (Crossgrove & Curran, 2008), the lack of a statistically significant difference in achievement between groups is a reasonably expected outcome.

Another important component of effective instruction deficient in other low-tech methods of traditional instruction, yet afforded by electronic response technology, is an efficient method of getting immediate feedback to the teacher (Judson & Sawada, 2002). Feedback regarding the class's level of conceptual understanding is extremely important (Black & Wiliam, 1998a, 1998b; Poulis et al., 1998) for effective teaching. With efficient methods for formatively assessing students' conceptual understanding, teachers can generate immediate feedback in ways that make daily lesson plans, even distinct lectures, dynamic in response to the background knowledge and misconceptions students bring to class or the pace of knowledge acquisition within a class. Given the ease with which a teacher can formatively assess a class and then gather, aggregate, and display results with electronic response technology, it is reasonable to expect such an advantage to manifest in higher performance for students in classes where such technology is used. However, in this study the teacher saw the Treatment group in the class period preceding the Comparison group on a daily basis. A confounding variable to the study design is that, in critical ways, the class sequence may have controlled for the advantage of efficient and immediate feedback afforded by electronic response technology. In other words, the students in the comparison group may have benefited from an order effect.

Because the teacher had feedback from daily formative assessments from students in the Treatment group, she was more aware of potential misconceptions and had greater capacity to recognize concepts likely to be difficult for the Comparison group. It is not possible from this design to consider whether such insight influenced instruction in positive ways, thereby increasing both instructional effectiveness and student learning for students in the Comparison group. It is entirely possible that the electronic response technology used by the Treatment group resulted in enhanced instruction for students of the Comparison group. In hindsight, the preferred treatment sequence probably would have been to instruct the Comparison group each day prior to the Treatment group. At least to a degree, such a sequence might have addressed this confounding variable.

If decisions were based only on the findings of the first question, I could not support the use of clicker technology given the lack of significant differences between groups. However, when one views the use of clicker technology through the lens of student engagement, the technology seems to hold promise.

Differences Between Groups on Levels of Active Engagement Across the Class Period

A secondary question focused on the difference between the Treatment and Comparison groups' mean levels of active engagement across the entire class period. Findings indicated statistical significance, $p < .01$, with an accompanying large effect size, $d = 1.28$. This large effect size indicated the educational significance of these differences. The data showed students in classes where clicker technology was a consistent part of the instructional routine were engaged at levels significantly higher than students in classes where clickers were not used. Given that student engagement is regarded as one of the better predictors of learning (Carini et al., 2006; Newmann, 1992), it is reasonable to assume that students, engaged at such high levels, would demonstrate significantly higher learning gains. Data from this study do not support this hypothesis.

Findings from this study are of particular importance not only because of their statistical significance but also because of the method of data collection. Unlike previous studies relying on student surveys or interviews (Guthrie & Carlin, 2004; Newmann, 1992; Paschal, 2002; Poulis et al., 1998; Suchman et al., 2006), surveys of institutional practices (Carini et al., 2006; Kuh, 2001) or anecdotal information from the teacher's perspective (Brown, 1972; Caldwell, 2007; Littauer, 1972; Wood, 2004), engagement data for this study came from direct observation of student behavior across an entire class period. Observed behavior was recorded as falling into one of three categories: (a) on-task, (b) passive engagement, or (c) off-task. The data presented in this study reflect how students *actually* behaved in class as opposed to how they *perceived* they behaved. Admittedly, there was some degree of subjectivity in that true cognitive engagement is intangible with observations relying on identifiable proxies such as attention and problem solving (National Research Council and the Institute of Medicine, 2004). While a potential confound existed due to the subjectivity of the observer, the confound was largely controlled, as the same skilled person collected data from both groups throughout the entire study using the same schematic to evaluate student behavior. Consequently, no bias should have been present.

Kuh (2001) suggested that what students do is the most important factor in determining their level of engagement. This perspective is well supported by research on the impact of active learning strategies (Duncan, 2005; Chickering & Gamson, 1987; Freeman et al., 2007; Hake, 1998; Marrs & Novak, 2004; Smith et al., 2005) asserting that students who are actively involved in the learning process are more engaged in the content and more likely to learn. Because technology motivates students to learn (Fox, 1988; Heafner, 2004; PIALP, 2008; Warschauer, 1996) and because clicker technology, specifically, affords opportunities for active learning (Guthrie & Carlin, 2004; Hake, 1998; Judson & Sawada, 2002), it is reasonable to assume that the use of clickers will consistently result in greater student engagement. Such reasoning has proven true in this study. Specifically, when compared to traditional methods of instruction, using clickers represents an effective, active learning strategy resulting in significantly higher levels of student engagement.

I built the theoretical-empirical framework for this study on the notion that increased student engagement would lead to increased academic achievement. However, there was not necessarily a clear relationship between the two in this study. Although students in the Treatment group were significantly more engaged than students in the Comparison group, the groups' academic achievement was not statistically different. I presented one possible explanation in my discussion of the first research question: namely that, because of the small sample size, the study could have lacked the statistical power to detect significance related to academic performance. Given the difference in levels of engagement and the large effect size, $d = 1.28$, one could reasonably have expected a significant difference in academic performance between the groups, but, again, that was not the case. A few thoughts come to mind when trying to explain the findings in this area.

The first thought is of the possibility of an engagement threshold. Namely, is there a point at which a student is adequately engaged enough to learn but where increased levels of engagement alone do not result in proportionately increased learning? While research tells us that inadequate levels of engagement negatively impact student learning, no research has been conducted to address just how much engagement is enough. An important outcome of this study is that neither of the two groups fell into the category of low engagement, as evidenced by their significant growth over time. Even though the Treatment group was significantly more engaged, that should not be interpreted as saying the Comparison group was disengaged. It is possible that both the Treatment and the Comparison groups were engaged beyond the aforementioned threshold, where despite greater engagement by one group there was no difference in learning gains.

The second thought has to do with the degree to which engagement in the use of the technology itself taxes the cognitive resources available for learning. If students focus on using the technology as an end rather than a means to learning a given content area, then inadequate cognitive resources would remain that could be devoted to the learning of that content. Likewise, it is possible the simple act of using the technology requires an amount of cognitive resource such that the remaining resources are inadequate for learning.

It is important, however, to note that both the Treatment and Comparison groups showed significant growth over time. These findings support the research linking engagement to desirable learning outcomes (Carini et al., 2006; Eckert, 1989; Hake, 1998). Even though the Treatment group was statistically more engaged than the Comparison group, I do not suggest the Comparison group was necessarily disengaged. Ultimately, my findings support the notion that achievement results from multiple factors, of which engagement is one (Carini et al., 2006).

Beyond the realm of academic achievement, clickers do seem to hold promise for the classroom. By simple definition, if students, when in class where clickers are used, demonstrate more engagement, they are on task more often. Such was the case in this study—not only when actively using the clickers but across the entire class period as well. This has very positive implications for the classroom environment and makes clicker use that much more compelling. If students demonstrate fewer inappropriate, undesirable and potentially distracting behaviors, teachers are less burdened by demands related to classroom management and are better able to focus on effective instruction. Likewise, engaged students are more likely to have positive experiences in the class (Caldwell, 2007; Carini et al., 2006; Crossgrove & Curran, 2008) and are therefore more motivated to learn (Keller, 1987). The positive impacts of clicker use on engagement hold even more promise during teacher questioning.

Differences Between Groups on Levels of Active Engagement Across Teacher Questioning

The third question asked whether there would be a significant difference between the Treatment and Comparison groups' mean levels of active engagement during teacher questioning because of the Treatment group's use of the electronic response technology.

The findings showed statistical significance, $p < .01$, with a large effect size, $d = 1.08$, indicating the educational significance of differences in scores. Similar to my findings regarding engagement across a class period, students in the Treatment group demonstrated significantly higher levels of engagement across teacher questioning than did students in the Comparison group. This fact is, again, interesting given the lack of statistical significance in the differences between groups on posttest scores. However, the same points could be made when addressing this question as were made when addressing the last—namely, that the study potentially lacked the necessary statistical power to detect significant differences in academic performance, the possibility of a threshold effect or the possibility that the use of the technology itself resulted in inadequate cognitive resources available for learning.

Likewise, the same arguments hold true about the promise of this technology. Increased student engagement across teacher questioning means fewer students were off task, demonstrating undesirable and potentially distracting behaviors during the most critical parts of classroom instruction. Thalheimer (2003) stated that students are more likely to learn when investing themselves enough to provide an answer to a question, even when only guessing. When well written, the questions teachers ask of students have as much or more to do with knowledge acquisition as any instructional technology (Beatty et. al., 2006; Bullock et. al., 2002; Reay et. al., 2008). If students are more engaged during teacher questioning because of clicker technology, as was the case in this study, then students using clickers are more likely to learn at high levels. This point alone is valuable and well worth the financial obligations associated with implementing this instructional strategy.

Full appreciation of the role electronic response technology plays in instruction during teacher questioning required a closer look at observed student behaviors in the Treatment and Comparison groups. Unlike the Comparison group, where students volunteered answers to teacher questions by raising their hands, all students in the Treatment group answered every question posed by the teacher. This is an important distinction. Even though students from the Comparison group were not overtly disengaged, they could choose whether or not to engage in the question cognitively. Because students are familiar with technology and have a positive feeling about its use (PIALP, 2001, 2008), they are likely to have the confidence to answer questions asked by the teacher when technology facilitates the exchange (Warschauer, 1996), and they are more motivated to do so (Fox, 1988; Heafner, 2004). Consequently, students are more likely to benefit from teacher questioning (Thalheimer, 2003) in classes where electronic response technology is a consistent part of the instructional routine.

Given all things considered in this study, I am confident electronic response technology is efficacious in ninth-grade science classrooms. I base my conclusion on the fact that electronic response technology enables teachers to efficiently utilize frequent formative assessments and make instructional adjustments based on immediate feedback about all students in the class in ways other, low-tech methods do not allow. Additionally, electronic response technology results in increased levels of student engagement during instruction. Finally, students in classes using the technology demonstrated statistically significant growth. Although their growth was not significantly different from students who did not use the technology, I feel the order in which the teacher taught the classes ultimately served to benefit students in the Comparison group. Specifically, increased awareness of emerging conceptions and misconceptions as well as knowledge of the most effective instructional approach resulted from teaching the Treatment group before the Comparison group. The Comparison group's enhanced instruction masked the impact of the intervention. Beyond learning gains alone, the evidence collected in this study points to higher quality instruction, increased student engagement across a class and increased student engagement across teacher questioning.

Implications for Practice

As a public school administrator, I constantly look for ways to help teachers identify efficiencies and increase the effectiveness of their instructional approach. Electronic response technology has great promise in K-12 education because of its well-documented impact on student engagement and its potential for increasing academic achievement through efficient, formative assessment routines. Keeping in mind Draper's (1998) premise that technology is just a tool, effective only when addressing a deficiency, educators should be aware that electronic response technology motivates students to learn and enables immediate feedback in ways not inherent to traditional instructional methodologies.

A vast body of research highlights technology's positive impact on students' motivation to learn (Fox, 1988; Guthrie & Carlin, 2004; Heafner, 2004; Judson & Sawada, 2002; PIALP, 2008; Warschauer, 1996). Findings from this study support conclusions from similar research studies in that students using technology were more motivated to learn, as demonstrated by higher levels of engagement. Because technology grabs students' attention, and because students find the use of technology in the classroom relevant to their extracurricular lives, students are more motivated to engage in instruction. When using technology, students' focus shifts from the task itself to the medium used to complete the task. Allowed to go unchecked, this could be counterproductive to the learning process. When used appropriately, however, technology increases motivation to learn because students are familiar with technology and are therefore confident in their use of it when participating in classroom activities. Finally, students find the use of technology satisfying, evidenced by the unprecedented explosion in its use (PIALP, 2008), thereby increasing their satisfaction with the learning process. When students are motivated to learn, they are more engaged in classroom activities. Given the relative ease and low cost of implementing this technology, its positive implications for student motivation and engagement are great.

Beyond increasing motivation to learn, clicker has even greater implications for enhancing teachers' ability to formatively assess student learning. According to Black and Wiliam (1998a, 1998b), formative assessment is the most effective innovation ever studied in the field of education. Formative assessment is a systematic process to consistently gather data regarding students' knowledge acquisition and disseminate information about learning while instruction is ongoing (Black & Wiliam, 1998a, 1998b). For an assessment to be formative, instructional adjustments must be made based on the data collected in an attempt to close the gap between knowledge acquisition and prescribed learning outcomes (Wiliam &Thompson, 2007). Although teachers can, and currently do, infuse their instructional routine with formative assessment, electronic response technology increases the efficiency of assessment administration as well as the immediacy of feedback to both teachers and students in ways not possible with low-tech or no-tech strategies.

Electronic response technology makes the practice of formative assessment easier and more efficient. With traditional question-and-answer methods to check for understanding, students volunteer answers by using one of two general methods: raising their hands or submitting answers in writing. When students have to raise their hands in response to a question, a number of variables serve as potential barriers to the teacher's acquisition of valid information about the level of content mastery in the class.

Rarely do all students raise their hands in an attempt to answer a question posed by the teacher. Because of this reluctance, students are relegated to a more passive role in discussions where they can choose either to engage or disengage in the learning process. Even if all students did raise their hands, the teacher could call on only one student at a time. Unless a teacher were to develop a set of questions large enough for everyone in the class and vetted to ensure all items sampled essentially the same concept, it would be impossible to formatively assess *all* students using this method. Beyond that issue, teachers would still face challenges from students unwilling to answer a question from fear of public humiliation in a process that could take hours. Such practices would quickly, if not immediately, become untenable. In addition, this problem would increase as class size increased.

While a teacher may avoid some of these issues by selecting a written format for student responses, the instructor's ability to ensure that all students answer all questions is limited. It is common for some students to refrain from answering some of the questions either because they do not know the answer or are not motivated to demonstrate their knowledge in such a format. Either way, it is difficult in such circumstances to make valid inferences about students' knowledge acquisition.

With clickers, a teacher can administer a small yet sophisticated formative assessment consisting of 5 to 10 questions to a class, thus allowing *all students* to respond in ways that are motivating rather than intimidating. Although there are time requirements associated with developing the items for the assessment instrument, the problems associated with a time requirement are not unique to electronic response systems. Regardless, recent research (Beatty et. al., 2006) suggests that more teacher preparation time needs to be devoted to developing high-quality, diagnostic questions than current practice reflects.

Aside from more effective and efficient administration of formative assessments, the immediacy of the feedback allowed by electronic response technology far exceeds the capacity of traditional methods of instruction. With clickers, teachers immediately have the benefit of aggregated and disaggregated responses. When used at the beginning of a unit, teachers could quickly ascertain the level of understanding students bring with them. Consequently, a teacher is able to align instruction to the current level of understanding and avoid waste precious instructional time by teaching something students already know. When clickers are used in the middle of a unit, teachers can evaluate the degree to which students are mastering the content. Because feedback is immediate, teachers can change the pace of instruction within a unit, or even a lesson. With immediate feedback, teachers can decide to cover certain materials again within the same lesson. When clickers are used to pose diagnostic questions, a teacher can use information from the responses to address emerging misconceptions within a single class period.

While teacher feedback is immediate when students use clickers, the same cannot be said for low-tech or no-tech methods of traditional instruction. Feedback regarding student progress is an essential component for effective teaching. Likewise, student feedback is an essential component of the learning process (Black & Wiliam, 1998a, 1998b).

With contemporary electronic response systems, aggregated class-wide responses are available immediately for students in the same way they are available to the teacher. In this way, students are able to evaluate their own learning in a less intimidating and interactive format. Students are able to compare the answers they gave to the answers submitted by other students in the class. When given the opportunity, students can discuss and defend their answers with their peers in the class—even before the teacher gives the correct answer. Because students can answer, and evaluate their answers, in real time, the connection between learning and teaching is strengthened.

In summary, although formative assessment is possible to some degree without any technology, electronic response systems allow for more effective and efficient assessment while providing immediate feedback not available with other methods to date. Given the user-friendly, low-cost and high student buy-in qualities of clicker implementation, the implications for practice are great.

Future Research

Increased use of technology within K-12 classrooms is inevitable. Each year, as student use of technology grows, researchers and practitioners alike are learning more about the factors effecting student motivation to learn and the mechanics driving increases in learning. Both policy and practice in education rely on findings from rigorous studies designed to add to a rapidly growing body of research. This study was a step in that direction. While much research is available on the use of electronic response technology in higher education settings, relatively little research is available on its effectiveness in 6th- to 12th-grade educational environments. Given the rate at which this technology is being purchased and used in secondary schools, a great deal more research is needed in order to ensure sound policy and instructional practice. To that end, there are several components of this study deserving future research.

The first suggestion would be to conduct a study involving a greater number of participants. Studies with greater numbers of participants may offer more power. Additionally, a larger group of study participants could reflect a more representative sample, resulting in findings generalizable to a greater number of students and schools. Given the increasingly rigorous graduation requirements adopted by nearly every state in the country, identifying factors positively influencing student achievement becomes more important every day.

Next, I recommend conducting this, or a similar, study but changing the sequence of the Treatment and Comparison groups' classes. I made the argument that instruction for the Comparison group was enhanced by virtue of the teacher using the technology in the Treatment group just before teaching the Comparison group, with both groups consisting of very similar students. Although this is a reasonable argument, it deserves the scrutiny afforded by a study with a rigorous design model.

Retention of knowledge is an important concept studied in higher education settings (Crossgrove & Curran, 2008). I suggest research-design methodologies focusing on the role electronic response technology plays in knowledge retention over time within K-12, and particularly 6th- to 12th-grade, settings. A research model similar to the one used in this study would be sufficient. In such a study, a posttest administered at the end of the two-term sequence required by the Science Department at this school (or even at the end of the academic year) could be analyzed to determine what, if any, difference in knowledge retention exists between the Treatment and Comparison groups. Coupled with findings from the study just completed, studies exploring the impact of clickers on knowledge retention over time would provide useful information when developing policy and making decisions related to instruction and technology integration.

As has been the case in higher education settings, I see the need for subsequent research exploring the impact of electronic response technology in content areas outside the sciences.

Finally, I would like to work with university researchers to better understand why the increased levels of engagement found in the Treatment group did not result in significantly higher levels of learning when compared to the Comparison group. Of particular interest would be designing a study to look for the possible existence of an engagement threshold as described previously in the study. Findings from this study take to task the notion that increased engagement results in increased learning. A better understanding of the causal mechanisms underlying the role of engagement in learning would benefit both research and practice.

APPENDIX

TEST INSTRUMENT

Science 9 A Pre-Test – Please do not write on this.

Science Skills

1. The metal is grey and shiny. Is this an observation or inference?
   a. Observation
   b. Inference

2. I think it is going to snow. Is this an observation or inference?
   a. Observation
   b. Inference

3. In a lab report, the analysis section is where you restate the purpose, summarize what happened in the lab, and state any errors that may have affected the results.
   a. True
   b. False

4. The liquid in the container looks like vegetable oil. Is this an observation or inference?
   c. Observation
   d. Inference

5. An experiment was done to see if the amount of fertilizer had an effect on how tall plants grew. The same type of plant and soil were used. In this experiment, what is the independent variable?
   a. The amount of fertilizer
   b. How tall the plants grew
   c. Type of plant
   d. Type of soil

Units – Match the correct units.

______ 6. Equilibrium	a. m/s^2
______ 7. Unit of Force	b. When the net force = 0
______ 8. Unit of Velocity	c. Newton
______ 9. Unit of Acceleration	d. m/s
______ 10. Inertia	a. kg
______ 11. Unit of Momentum	b. g-m/s
______ 12. Unit of Mass	c. Resisting change in motion
______ 13. Unit of Weight	d. Pounds

Motion/Speed/Velocity

14. What is it called when an object changes position relative to something else?
    a. Density b. Distance c. Motion d. Inertia

15. Speed is calculated by taking distance divided by ________.
    a. Time b. Mass c. Acceleration d. Force

16. A bicyclist travels 50 miles in 2 hours. What is the average speed of the bicyclist?
    a. 100 mph b. 25 mph c. 25 m/s d. 100 m/s

17. What is the velocity of a motorcycle that travels west 3,000 meters in 10 seconds?
    a. 30 m/s, west
    b. 300 m/s, west
    c. 300 m/s
    d. 30,000 m/s, west

18. Kiki was driving down the highway at 70 mph, N. This is BEST described as a:
    a. Speed
    b. Velocity
    c. Acceleration
    d. Momentum

Acceleration

19. In which of the following conditions does the car NOT accelerate?
    a. A car moves at 80 km/hr on a flat, straight highway
    b. The car turns a corner
    c. The car slows from 80 km/hr to 35 km/hr
    d. The car speeds up from 35 km/hr to 80 km/hr

20. A car coming to a stop sign changed velocity from 20 m/s to 0 m/s in 4 seconds. What was the acceleration of the car?
    a. 5 m/s^2
    b. 80 m/s^2
    c. –5 m/s^2
    d. 0 m/s^2

21. Which of the following objects has NEGATIVE acceleration?
    a. A jogger moving at a constant speed
    b. A car that is slowing down
    c. Earth orbiting the sun
    d. A car that is speeding up

22. Which of the following can occur when an object is accelerating?
    a. It speeds up
    b. It slows down
    c. It changes direction
    d. All of the above

23. A ball dropped from the top of a cliff has a positive acceleration. The ball must be ______________.
    a. Speeding up.
    b. Slowing down.
    c. Staying the same speed.

Momentum

24. Momentum of an object depends on what two things?
    a. Distance and time
    b. Speed and velocity
    c. Mass and velocity
    d. Mass and acceleration

25. A large boulder and a small pebble are rolling down a hill at the same speed. Which has more momentum?
    a. Large boulder
    b. Small pebble
    c. They both have the same amount of momentum.

26. What is the momentum of a 1.5 kg ball flying at 13 m/s?
    a. 19.5 kg-m/s b. 8.66 kg-m/s c. 14.5 kg-m/s

27. If you increase the mass of an object that is moving, the momentum will increase.
    a. True
    b. False

Force and Motion

28. Force is calculated by taking mass times:
    a. Time b. Mass c. Acceleration d. Velocity

29. Forces always occur in:
    a. Units b. Pairs c. Space d. Mass

30. How much force is required to move a 15 kg wagon at an acceleration of 3.5 m/s^2?
    a. 4.3 N b. 4.3 kg-m/s c. 52.5 N d. 52.5 kg-m/s

31. Which term below best describes the forces on an object with a net force of zero?
    a. Inertia b. Balanced forces c. Acceleration d. Unbalanced forces

32. The forces acting on an object are 3 N east, 2 N south, 2 N north, and 4 N west. The net force is:
    a. 11 N west
    b. 0 N
    c. 1 N west
    d. 1 N east

33. If an object is accelerating, the forces acting on the object must:
    a. Be in equilibrium
    b. Have a net force of 0 N.
    c. Be balanced
    d. Be unbalanced.

Newton's Laws of Motion

34. Which of the following is an example of Newton's Third Law of Motion?
    a. A book on the desk stays there until someone pushes it.
    b. A fish using its fins to push on water, which then pushes back on its fins.
    c. A wagon being pulled on by the handle will accelerate in the direction of the pull.
    d. None of the above.

35. If you give a skateboard a push, it will eventually come to a stop. The force that is MOST responsible for the stop is:
    a. Tension Force
    b. Normal Force
    c. Applied Force
    d. Friction Force

36. You pull a piece of paper very quickly out from underneath a glass. The glass stays in the same spot. This is an example of:
    a. Newton's 1st Law
    b. Newton's 2nd Law
    c. Newton's 3rd Law

37. When you pull on the handle of a wagon, you are causing it to accelerate in the direction of the force of your pull. This is an example of:
    a. Newton's 1st Law
    b. Newton's 2nd Law
    c. Newton's 3rd Law

38. Which of Newton's Laws explains why a person who is wearing a seat belt will most likely not hit the windshield in the event of a collision?
    a. Newton's 1st
    b. Newton's 2nd
    c. Newton's 3rd
    d. Not enough information

Work and Machines

In the following statements, decide if Work or No Work was done:

39. A pulley was used to lift a crate to the second floor.
    a. Work
    b. Not Work

40. A textbook was pushed off of a desk and fell to the floor.
    a. Work
    b. Not Work

41. A girl held a rock in her hand.
    a. Work
    b. Not Work

42. I raked all the leaves in my backyard.
    a. Work
    b. Not Work

43. A textbook was sitting on a desk.
    a. Work
    b. Not Work

44. What do we call the rate at which work is done?
    a. Efficiency
    b. Work
    c. Force
    d. Power

45. What can be said about a machine where the work input closely matches the work output?
    a. It is efficient
    b. It is inefficient
    c. It is fast.
    d. It is slow.

46. In which situation was work done?
    a. A person pushed against the wall and it did not move.
    b. A book was sitting on a desk.
    c. A box was moved from one side of the room to the other.
    d. All of the above.

47. What are the units for work?
    a. Watts
    b. Newtons
    c. Joules
    d. Meters

48. What class of lever is a wheelbarrow an example of?
    a. A first class lever
    b. A second class lever
    c. A third class lever
    d. A pulley

Potential/Kinetic Energy

Decide whether each example below is Potential Energy or Kinetic Energy:

a = PE b = KE

49. A moving skateboard
50. A rock at the edge of a cliff
51. A glass of milk
52. Gasoline
53. A basketball passing through a hoop
54. A battery
55. Blowing wind
56. A book on the table

Heat and Temperature

57. The Kelvin temperature scale is based on the freezing and boiling points of water.
    a. True
    b. False

58. What unit is heat measured in?
    a. Joules
    b. Watts
    c. Calories
    d. Degrees

59. Heat energy flows from warmer to cooler materials.
    a. True
    b. False

60. Which temperature scale is based on the concept of absolute zero?
    a. Kelvin
    b. Celsius
    c. Fahrenheit
    d. None of the above

REFERENCES

Bapst, J. J. (1971). *The effect of systematic student response upon teaching behavior*. Unpublished doctoral dissertation, University of Washington, Seattle. (ERIC Document Reproduction Service No. ED060651)

Barber, M., & Njus, D. (2007). Clicker evolution: Seeking intelligent design. *Cell Biology Education, 6*, 1–8.

Beatty, I. D. (2004). *Transforming student learning with classroom communication systems*. Retrieved March 08, 2008, from http://www.utexas.edu/academic/cit/services/cps/ECARCRS.pdf

Beatty, I. D., Gerace, W. J., Leonard, W. J., & Dufresne, R. J. (2006). Designing effective questions for classroom response system teaching. *American Journal of Physics, 74*(1), 31–39.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*, 536–553.

Bessler, W. C., & Nisbet, J. J. (1971). The use of an electronic response system in teaching biology. *Science Education, 3*, 275–284.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5,* 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*, 139–147.

Bonwell, C., & Eison, J. (1991). *Active learning: Creating excitement in the classroom (AEHE-ERIC Higher Education Report No. 1)*. Washington, DC: Jossey-Bass.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems. *Annals of Statistics, 25*, 290–302.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design and data analysis*. New York: John Wiley.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience and school.* Washington, DC: National Academy Press.

Brown, J. D. (1972). An evaluation of the Spitz student response system in teaching a course in logical and mathematical concepts. *Journal of Experimental Education, 40*(3), 12–20.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*(1), 21–32.

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*, 30–36.

Bunce, D. M., Van den Plas, J. R., & Havanki, K. L. (2006). Comparing the effectiveness on student achievement of a student response system versus online WebCT quizzes. *Journal of Chemical Education, 83*, 488–493.

Burnstein, R. A., & Lederman, L. M. (2003). Comparisons of different commercial wireless keypad systems. *The Physics Teacher, 41*, 272–275.

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

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education, 47*(1), 1–32.

Casanova, J. (1971). An instructional experiment in organic chemistry, the use of a student response system. *Journal of Chemical Education, 48*(7), 453–455.

Chickering, A. W., & Gamson, Z. (1987). Seven principles for good practices in undergraduate education. *AAHE Bulletin, 39*, 3–7.

Chickering, A. W., & Reisser, L. (1993). *Education and identity* (2nd ed.). San Francisco: Jossey-Bass.

Chu, Y. (1972). *Study and evaluation of the student response system in undergraduate instruction at Skidmore College*. Report prepared for the National Science Foundation, Arlington, VA. (ERIC Document Reproduction Service No. ED076135).

Corporation for Public Broadcasting. (2002). *Connected to the future*. Retrieved on February 21, 2008, from http://www.cpb.org/stations/reports/connected/

Crossgrove, K., & Curran, K. L. (2008). Using clickers in nonmajors- and majors-level biology courses: Student opinion, learning, and long-term retention of course material. *Cell Biology Education—Life Sciences Education, 7*, 146–154.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics, 69,* 970–977.

Draper, S. W. (1998). Niche-based success in computer assisted learning. *Computer Education, 30*, 5–8.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20*, 81–94.

Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L. (1996). Classtalk: A classroom communication system for active learning. *Journal of Computers in Higher Education, 7*, 3–47.

Duncan, D. (2005). Clickers in the classroom: How to enhance science teaching using classroom response systems, New York: Addison-Wesley and Benjamin Cummings.

Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school.* New York: Teachers College, Columbia University.

Fies, C., & Marshall, J. (2006). Classroom response systems: A review of the literature. *Journal of Science Education and Technology, 15*(1), 101–109.

Fox, M. (1988). A report on studies of motivation teaching and small group interaction with special reference to computers and to the teaching and learning of arithmetic. Milton Keynes, England: The Open University, Institute of Educational Technology.

Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., et al. (2007). Prescribed active learning increases performance in introductory biology. *CBE Life Sciences Education, 6*, 132–139.

Friedman, T. L. (2005). *The world is flat*. New York: Farrar, Straus, & Giroux.

Garg, D. P. (1975). *Experiments with a computerized response system: A favorable experience*. Paper presented at the Conference on Computers in the Undergraduate Curricula, Fort Worth, TX. (ERIC Document Reproduction Service No. ED 111355).

Garson, G. D. (2009). Univariate GLM, ANOVA, and ANCOVA. Retrieved on May 23, 2009, from http://faculty.chass.ncsu.edu/garson/PA765/anova.htm

Guthrie, J. T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behavior, 10*, 23–28.

Guthrie, R. W., & Carlin, A. (2004). *Waking the dead: Using interactive technology to engage passive listeners in the classroom*. Proceedings of the Tenth Americas Conference on Information Systems, New York.

Hake, R. R. (1998). Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66,* 64–74.

Hake, R. R. (2002). Lessons from the physics education reform effort. *Conservation Ecology, 5*, 28.

Halpin, G., Carwile, R., & Halpin, G. (1991, November 12-15). *So you have chosen an unequal cell size ANOVA option. Do you really know what you have?* Paper presented at Mid-South Educational Research Association, Lexington, KY.

Hayes, R. L., Nelson, J. L., Tabin, M., Pearson, G., & Worthy, C. (2002). Using school-wide data to advocate for student success. *Professional School Counseling, 6*, 86–94.

Heafner, T. (2004). Using technology to motivate students to learn social studies. *Contemporary Issues in Technology and Teacher Education, 4*(1), 42–53.

Jacobson, T. E., & Xu, L. (2002). Motivating students in credit-based information literacy courses: Theories and practice. *Libraries and the Academy, 2*(3), 423–441.

Judson, E., & Sawada, D. (2002). Learning from the past and present: Electronic response systems in college lecture halls. *Journal of Computers in Mathematics and Science Teaching, 21*(2), 167–181.

Keller, J. M. (1987). Development and use of the ARCS model of motivational design. *Journal of Instructional Development, 10*(3), 2–10.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education, 2,* 155–161.

Kuh, G. (2001). *National Survey of Student Engagement*. Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Littauer, R. (1972). Instructional implications of a low-cost electronic student response system. *Educational Technology: Teacher and Technology Supplement, 12*(10), 69–71.

Lowman, J. (1990, Fall). Promoting motivation and learning. *College Teaching, 38*, 137.

Marrs, K. A., & Novak, G. (2004). Just-in-time-teaching in biology: Creating an active learner classroom using the Internet. *Cell Biology Education, 3*, 49–61.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*(1), 14–19.

Mazur, E. (1997). *Peer instruction, a user's manual*. Upper Saddle River, NJ: Prentice-Hall.

McGregor, J. H. (1999). How do we learn? In B. K. Stripling (Ed.), *Learning and libraries in an information age: Principles and practice* (pp. 25-53). Englewood, CO: Libraries Unlimited.

McNeil, L. M. (1986). *Contradictions of control: School structure and school knowledge*. New York: Routledge.

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

National Research Council and the Institute of Medicine, Committee on Increasing High School Students' Engagement and Motivation to Learn. (2004). *Engaging schools: Fostering high school students' motivation to learn.* Washington, DC: The National Academy Press.

Newmann, F. M. (1992). *Student engagement and achievement in american secondary schools.* New York: Teachers College Press.

Nichol, D. J., & Boyle, J. T. (2003). Peer instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom. *Studies in Higher Education, 28*, 457–473.

Novak, G., Patterson, E. T., Gavrin, A. D., & Christian, W. (1999). Just-in-time teaching: Blending active learning with web technology, Upper Saddle River, NJ: Prentice-Hall.

Oregon Department of Education. (2007). Report cards. Retrieved on May 23, 2008, from http://www.ode.state.or.us/data/reportcard/RCpdfs/08/08-ReportCard-561.pdf

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco: Jossey-Bass.

Paschal, C. B. (2002). Formative assessment in physiology teaching using a wireless classroom communication system. *Advances in Physiology Education, 26*, 299–308.

Penuel, W. R., Roschelle, J., Crawford, V., Shechtman, N., & Abrahamson, L. (2004). *Workshop report: Advancing research on the transformative potential of interactive pedagogies and classroom networks*. Retrieved on February 21, 2008, from http://ctl.sri.com/publications/ Downloads/CATAALYST_ Workshop_Report.pdf

Penuel, W. R., Roschelle, J., & DeBarger, A. H. (2006). *Teaching with student response system technology: A survey of K-12 teachers*. Retrieved on February 21, 2008, from http://ctl.sri.com/publications/downloads/Teaching_with_ Audience_Response_Systems_Brief_Report.pdf

Pew Internet & American Life Project. (2001). *Parents, kids, and the Internet.* Retrieved on February 26, 2008, from http://www.pewinternet.org/pdfs/PIP_ Teens_Questionnaire.pdf

Pew Internet & American Life Project. (2008). *Surviving in the new digital ecosystem.* Retrieved on February 26, 2008, from http://www.pewinternet.org/ PPF/r/113/presentation_display.asp

Poulis, J., Massen, C., Robens, E., & Gilbert, M. (1998). Physics lecturing with audience paced feedback. *American Journal of Physics, 66*, 439–441.

Preszler, R. W., Dawe, A., Shuster, C. B., & Shuster, M. (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE Life Sciences Education, 6,* 29–41.

Rao, S. P., & DiCarlo, S. E. (2000). Peer instruction improves performance on quizzes. *Advances in Physiology Education, 24*, 51–55.

Reay, N. W., Li, P., & Bao, L. (2008). Testing a new voting machine question methodology. *American Journal of Physics, 76*(2), 171–178.

Roschelle, J., Penuel, W. R., & Abrahamson, L. (2004). *Classroom response and communication systems: Research review and theory.* Paper presented at the AERA conference. Retrieved on Februar 21, 2008, from http://ubiqcomputing.org/CATAALYST_AERA_Proposal.pdf

Smith, A. C., Stewart, R., Shields, P., Hayes-Klosteridis, J., Robinson, P., & Yuan, R. (2005). Introductory biology courses: A framework to support active learning in large enrollment introductory science courses. *Cell Biology Education, 4*(2), 143–156.

Sotto, E. (1994). *When teaching becomes learning: A theory and practice of teaching.* London: Cassell.

Suchman, E., Uchiyama, K., Smith, R., & Bender, K. (2006). Evaluating the impact of a classroom response system in a microbiology course. *Microbiology Education, 7*, 3–11.

Tallal, P., Merzenich, M., Jenkins, W. M., & Miller, S. L. (1999). Moving research from the laboratory to clinics and classrooms. In D. D. Duane & M. D. (Eds.), *Reading and attention disorders: Neurobiological correlates* (pp. 93–112). Baltimore: York Press.

Thalheimer, W. (2003, January). *The learning benefits of questions*. Retrieved November 22, 2008, from http://www.work-learning.com/ma/PP_WP003.asp

Udovic, D., Morris, D., Dickman, A., Postlethwait, J., & Wetherwax, P. (2002). Workshop biology: Demonstrating the effectiveness of active learning in an introductory biology class. *Bioscience, 52*, 272–281.

United States Department of Education. (2004). *Visions 2020.2: Student views on transforming education and training through advanced technologies.* Retrieved on February 26, 2008, from http://www.nctet.org/Visions2020-2.pdf

Watkins, C. L., & Slocum, T. A. (2003). The components of direct instruction. *Journal of Direct Instruction, 3*, 75–110.

Warschauer, M. (1996). Motivational aspects of using computers for writing and communication. In Mark Warschauer (Ed.), *Telecollaboration in foreign language learning: Proceedings of the Hawaii symposium* (Technical Report No. 12, pp. 29–46). Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center. Retrieved February, 1, 2009, from http://www.lll.hawaii.edu/nflrc/NetWorks/NW1/

Weis, L. (1990). *Working class without work: High school students in a decentralized economy.* New York: Routledge.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping, teaching, and learning* (pp. 365–414). Mahwah, NJ: Lawrence Erlbaum.

Wit, E. (2003). Who wants to be. . . . The use of a personal response system in statistics teaching. *MSOR Connections, 3*(2), 14–20.

Wood, W. B. (2004). Clickers: A teaching gimmick that works. *Developmental Cell, 7*(6), 796–798.