Implementing a Mastery Model through Self Quizzing in an Online Learning Environment

Amy E. Scrima

Western Michigan University

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

Part of the Cognitive Psychology Commons, Educational Psychology Commons, and the School Psychology Commons

Recommended Citation
Scrima, Amy E., "Implementing a Mastery Model through Self Quizzing in an Online Learning Environment" (2009). Dissertations. 694.
https://scholarworks.wmich.edu/dissertations/694
IMPLEMENTING A MASTERY MODEL THROUGH SELF QUIZZING IN AN ONLINE LEARNING ENVIRONMENT

by

Amy E. Scrima

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology
Advisor: Richard Malott, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2009
IMPLEMENTING A MASTERY MODEL THROUGH SELF QUIZZING IN AN ONLINE LEARNING ENVIRONMENT

Amy E. Scrima, Ph.D.

Western Michigan University, 2009

Mastery learning has an extensive and long-standing research base as an efficacious instructional methodology. The use of mastery learning with current technological advances, however, is a new endeavor. The current study evaluated the effects of adding a mastery learning component to an introductory college course by using an online course management system to facilitate frequent, self-given, chapter review quizzes. Thirty-two first- and second-year college students of similar demographic makeup at a midsize community college were the participants for this study. An alternating treatment design was used to assign students to mastery and non-mastery conditions in two sections of an introductory psychology course. The effectiveness of this strategy was measured by assessing the differences between the treatment conditions using unit and final exam scores compared to the counterbalanced control conditions. No apparent differences between treatment conditions were found. Social validity data were collected and are discussed in terms of student attitudes and instructor usability for this approach. Possible explanations for the outcomes as well as suggestions for future research directions in this area are discussed.
Copyright by
Amy E. Scrima
2009
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction. In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
ACKNOWLEDGMENTS

I would like to express my most sincere gratitude to several individuals without whom this endeavor would not have been possible. First of all, I would like to thank my committee members and many advisors through the years. To Dr. Dick Malott, my academic advisor and philosophical mentor, thank you for showing me the path to truth. To Dr. Dale Brethower, also my advisor and mentor, thank you for introducing me to Behavior Analysis and, particularly, to Human Competence. My life was changed forever that winter of 1997 when I started graduate school by your powerful words and exemplary example. Thank you also to Dr. Eric Fox, who is responsible for many of the ideas that led to the creation of this study. This dissertation would not be what it is without your thoughtful and technically meticulous help. I have become a better writer and researcher for your tireless effort. And to Dr. Kristal Ehrhardt, thank you so much for your time and willingness to serve on my committee. While rather unorthodox, I highly recommend having more than one advisor; I have learned so much from each of them and each has made me a better behavior analyst and better person.

Next, I would like to acknowledge and thank my colleagues at Lake Michigan College, where this study took place. First, to my department chair, Denise Scameheorn, who rearranged schedules more than once on my behalf, often to take on extra classes herself so that I could keep working, thank you so much for your support. To Mark Kelly and Donna Maxon, the computer geniuses who worked tirelessly to learn the ins and outs
of Blackboard along with me—thank you. And my most heartfelt appreciation to Judy Buchalski, my friend and mentor, without whom I would not be here. Thank you for believing in me, even before there was anything to believe in.

To my fellow students, who saw me through the comps, Kristen Gaisford, Breanne Hartley, and Woan Tian Chow, thank you for your acceptance and invaluable help. Thank you to my research assistant and friend Megan Knight, who kindly and persistently helped me with whatever was needed. Thank you especially to my many amazing friends who watched my kids, listened to my complaining, brought me chocolate-covered pretzels and baked me bread, especially Beth Suarez and Marilu Garcia. I could not have done it without you.

My last acknowledgments have to be to my family. First, to my Dad and Mom, whose complete faith in me has allowed me to do things I would not have done, if I had ever stopped to question whether I could. Thank you to my sisters Alison and Darla and my brother Mark, who continue to inspire me to be the best I can. To my husband Brian, who has put up with a lot in the name of education for most of the 14 years we’ve been married, thank you for your patience and support, but most of all love. And to my children, Jessi and Parker, whose involuntary sacrifice these last two years has been the hardest part of this endeavor to endure, I will spend the rest of my life trying to make it up to you. Hopefully you will one day understand and appreciate this journey I’ve taken—and my greatest hope is that somehow, it may help you on your own.

Amy E. Scrima

iii
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ ii

INTRODUCTION ................................................................................................................... 1

Mastery Learning ........................................................................................................... 1

PSI Effectiveness .......................................................................................................... 6

Obstacles to Implementing PSI .............................................................................. 7

Modifications of PSI .................................................................................................. 10

Component Use Modifications .............................................................................. 11

Computer-Based Modifications .......................................................................... 21

Web-Based Technology in College-Level Instruction .......................................... 26

Instructor Adoption and Implementation .......................................................... 32

Rationale for Current Study ...................................................................................... 35

METHOD ................................................................................................................................. 36

Participants and Setting ............................................................................................. 36

Materials ......................................................................................................................... 37

Textbook ......................................................................................................................... 37

Course Website ............................................................................................................. 38

Mastery Quizzes ........................................................................................................... 38

Unit Exams ..................................................................................................................... 39

Final Exam ..................................................................................................................... 39
Table of Contents—Continued

Additional Points ................................................................................................. 39
Survey ................................................................................................................... 40
Experimental Design ............................................................................................. 41
Procedures ............................................................................................................. 41
Dependent Variables ............................................................................................ 44
Social Validity ......................................................................................................... 44
Instructor Usability ................................................................................................. 45
Data Collection and Analysis ............................................................................... 45
RESULTS ................................................................................................................ 45
Exam Scores ........................................................................................................... 45
Intergrader Reliability .......................................................................................... 47
Final Exam Scores ................................................................................................ 47
Mastery Quiz Scores ............................................................................................ 48
Mastery Quiz Attempt .......................................................................................... 48
Social Validity ......................................................................................................... 49
Instructor Usability Data ....................................................................................... 51
DISCUSSION .......................................................................................................... 51
REFERENCES ......................................................................................................... 64
APPENDICES
A. Human Subjects Institutional Review Board Letter of Approval .................... 71
B. Human Subjects Institutional Review Board Informed Consent Form ............ 73
APPENDICES

C. Example of Unit Exam ............................................................................................... 76
D. Example of Reading Quiz ........................................................................................... 78
E. Example of Assignment .............................................................................................. 80
F. Demographics Questionnaire ..................................................................................... 82
G. Social Validity Survey ................................................................................................ 84
H. Assignment Schedule .................................................................................................. 86
I. Sample IOA Calculation Sheet .................................................................................. 89
J. Graphs of Individual Student Averaged Exam Scores by Condition .................... 91
K. End of Semester Survey Student Responses to Question 7, “What if anything would you change about the review quizzes?” ........................................103
L. Tables and Figures ........................................................................................................106
INTRODUCTION

Mastery may or may not be a stated goal of instruction, but most instructors would say it should be. Whether instructors hope for it or arrange for it is a choice they can make. Helping instructors make the right choice, both for themselves and for their students, is a worthwhile educational aim. To achieve such a goal, two things must be established: how mastery can most efficiently and effectively be achieved and how such an arrangement can best be implemented. Several issues and areas of research seem particularly relevant to pursuing this goal in higher education: mastery learning and the Personalized System of Instruction (PSI; Keller, 1968), the use of web-based technology to support instruction, and instructor usability.

Mastery Learning

Mastery learning is an instructional methodology in which students are required to demonstrate proficiency in a learning objective before advancing to subsequent course content. Comprehension of the learning material is measured by frequent assessments, with a predetermined mastery criterion set for each. If the mastery criterion is not achieved on an assessment, the learning materials are restudied and another assessment is taken until the criterion is met. This process of study and assessment is repeated until all sections of the course have been mastered. There is some variation in the implementation of mastery learning programs and the structure of the methodology, but the goal of mastery remains constant.
There is a long and extensive literature base for mastery learning. Kulik, Kulik, and Bangert-Drowns (1990) summarized the empirical research on mastery learning programs completed prior to 1990 in a meta-analysis. Meta-analyses use a statistical method to determine the overall effect size of a given treatment across multiple studies (Glass, 1976), and the Kulik et al. analysis showed a significant treatment effect for courses with a mastery learning methodology as compared with traditional lecture-based courses. Kulik et al. examined 108 studies, and 103 of them reported end-of-course examination scores that were used to perform statistical analyses. The effect size found for these 103 mastery learning programs was 0.52. This means that, on average, mastery learning programs had the effect of raising scores by 0.52 standard deviations. This effect is also highly statistically significant, $t(102) = 15.78, p < .001$. Another way to view this data is that an average student in a mastery-based course scored in the 70th percentile, while an average student in a non-mastery (traditional) course scored in the 50th percentile.

The Kulik et al. (1990) meta-analysis focused on two main programs, both developed in the 1960s: Keller’s Personalized System of Instruction (PSI; Keller, 1968) and Bloom’s Learning for Mastery (LFM; Bloom, 1968). These programs are recognized as the most influential and widely used mastery learning programs. PSI and LFM are similar in their focus on mastery, but differ in their approach to helping students achieve it.

PSI focuses student attention on short sections of (traditionally) written text, directing students to take quizzes over the material when they feel they are ready. When a quiz is passed, usually with a 80-100% criterion of mastery (Kulik et al., 1990), the
student continues on to the next unit of the course. If a quiz is not passed, the student is
directed to restudy the material and take another quiz when he or she is prepared. This
process continues until all units in a course are completed to mastery. Traditionally, the
PSI program is facilitated by student proctors who administer and grade the unit quizzes
and are available for consultation throughout the course.

LFM uses group lessons, presented by an instructor. At the end of a lesson,
students take a quiz, usually with a mastery criterion set at 70-100% (Kulik et al., 1990).
Those students who pass the quiz go on to the next group lesson, while those who do not
are tutored, either individually or in a group, until the instructor deems the student(s)
ready to retake the quiz for that lesson. Like PSI, this process continues until each student
achieves mastery in all units of the course.

Bloom (1968) described what he thought were the most important outcomes of
any mastery-based course by asserting that, because instruction is the same for all
students, the end result of proficiency should also be the same. This led to his theory of
decreasing variability, which essentially states that 90% of students in any mastery-based
course should achieve the same level of proficiency, previously achieved by only the top
10% in lecture-based courses. This assertion has been questioned and, in some cases, has
not been supported by the literature (Arlin, 1984; Livingston & Gentile, 1996).

PSI and LFM are similar in their requirement of mastery to a specific criterion for
continuation through a course of study. As such, the stated goal for each approach is for
students to achieve mastery. The way each program advances that goal is quite different,
however. One of the main differences between PSI and LFM is where the impetus is
placed for progress through the course. In PSI, responsibility for assessment of readiness
falls to the individual learner, whereas in LFM, it is a more instructor-controlled process. PSI, in its original form, is completely individualized; a student may progress as quickly as he or she is able and is unaffected by the performance of other students in the course. LFM, by contrast, is a group arrangement in which all learners in a course are exposed to material concurrently and individual progress is restricted by the progress of the group.

Another difference between PSI and LFM is the way research has been conducted to demonstrate effectiveness of these mastery models. Kulik et al. (1990) found many more PSI studies (72) than LFM studies (36) that met their inclusion criteria of a methodologically sound, empirical approach to assessment. This discrepancy exists partially because less research overall has been done concerning LFM effectiveness and because some LFM studies were found by Kulik et al. to be less methodologically valid.

The results of Kulik et al.’s (1990) meta-analysis were first classified across 14 salient features, such as pacing, mastery level, and course duration. This breakdown shows a clearer distinction between the two programs. For example, 100% of all PSI studies used a formal demonstration of mastery (e.g., a written test), compared with 30% of LFM courses. Despite these differences, the overall achievement effects, which have most often been measured by end-of-course exam scores, are very similar for both programs. Of the 103 mastery-learning studies examined (both PSI and LFM), 95 revealed positive effects and 67 of those 95 reported those effects to be statistically significant. These effects were large and positive overall, but the size of effect did vary predictably according to certain features. Those features were (a) level of mastery required, (b) use of standardized versus instructor developed test, (c) course content, and (d) duration of the study. When these features are taken into account, the data show that
(a) high mastery standards were more effective than lower standards, (b) studies with instructor-developed tests had larger effect sizes than those that used standardized tests, (c) effects were larger for courses in the social sciences than courses in the natural sciences and mathematics, and (d) the duration of the study was not found to be significantly related to size of effect (Kulik et al., 1990).

The authors also performed multiple regression analyses on three other features they thought might be intercorrelated with the first four, even though these features were not originally identified as being related to the size of achievement effect in the first order analysis of variance. The characteristics were (a) requirement of a formal demonstration of mastery, (b) group versus individualized pacing, and (c) the amount of feedback given to experimental and control groups. The requirement of formal mastery demonstration did not show a significant effect difference. The group pacing studies did show larger effects than the individualized pacing studies, however, and whether feedback was given to the control group also made a difference in the effect size. When the control group received the same amount of feedback as the experimental group, effect sizes were smaller, but if the control group received less feedback than the experimental group, the effect sizes were larger (Kulik et al., 1990).

These data show that both PSI and LFM are effective. There are, however, questions of efficiency and relevance with regard to LFM. The remainder of this section will focus on PSI as opposed to LFM for three reasons. First, most of the research on PSI has been conducted in higher education settings, which is the educational level of interest for this discussion, whereas about half the research on LFM is dedicated to applications in primary and secondary education (Kulik, Kulik, & Cohen, 1979). Second, the group
arrangement of LFM can result in a "leveling effect." Slavin (1987) noted that, in LFM, higher-achieving students must progress at a slower pace than they otherwise could if instruction were individualized: "it is likely to require taking time away from high achievers to increase it for low achievers, a leveling process that would in its extreme form be repugnant to most educators" (p. 178). Others argue that the leveling effect is not unique to LFM, but agree that group-formatted educational practices are, by design, beset with this unfortunate result (Anderson & Burns, 1987). Third, as a group-based format, LFM may not have the potential to be implemented as efficiently as PSI. With the trend in higher education toward more individualized instruction, as is clearly the case in distance education (Bradford, Porciello, Balkon, & Backus, 2007), the PSI model is a much more likely candidate for meeting both present and future higher education needs (Fox, 2004; Fredrick & Hummel, 2004).

**PSI Effectiveness**

PSI has a large and varied base of empirical studies showing its greater efficacy over lecture-based teaching. This evidence has been well established by hundreds of studies over the last four decades, notably the Kulik et al. (1990) mastery meta-analysis described above and an earlier meta-analysis focusing solely on PSI effectiveness (Kulik et al., 1979).

The Kulik et al. (1979) analysis examined 75 PSI studies and found that PSI produced "superior student achievement, less variation in achievement, and higher student ratings in college courses" (p. 1). They also found a strengthened effect for retention scores. When students were tested several months after the class, they scored on
average 14 points higher on examinations than traditionally taught students. This is a strengthened effect because the average point difference measured at the end of the courses was 8 points. They also found no evidence for increased course withdrawal rates or student study time.

Many other reviews and analyses have been conducted that look either at PSI alone (Buskist, Cush, & DeGrandpre, 1991; Hursh, 1976; Kulik, Jaksa, & Kulik, 1978; Kulik, Kulik, & Cohen, 1980; Robin, 1976; Taveggia, 1977) or in conjunction with other educational systems (Dubin & Taveggia, 1968). These studies also concluded that PSI was more effective overall than lecture-based interventions.

While implementation and research on PSI has steadily declined since the 1970s (Sherman, 1992), research continues to demonstrate its efficacy in many different disciplines, including medicine (Benbassat & Baumal, 2007), physical education (Harrison et al., 1999), and reading instruction (Reid, 1997). Research also supports the use of PSI with many different populations, including individuals with developmental disabilities (Brothen, Wambach, & Hansen, 2002; Zencius, Davis, & Cuvo, 1990), individuals with physical disabilities (Mann & Elland, 2005), and the intellectually gifted (Matthews & Foster, 2006). Despite this impressive range of applicability and interest, there are some limitations on its implementation and some criticisms of PSI, especially in its original form.

**Obstacles to Implementing PSI**

There are several obstacles to the traditional implementation of PSI. One is the seeming incompatibility of PSI's mastery requirement within the structure of the
predominant semester/term schedule used by most colleges and universities (Buskist et al., 1991). For example, a student working through a typical PSI course may need significantly less time to complete the requirements than the usual 14-16 week semester. This likely possibility would threaten the very structure of higher educational institutions (Buskist et al., 1991). Questions such as how the college would charge for a course and how the college would keep track of individual student progress are not quickly or easily answered. Societal issues such as the (young) age at which some students might graduate, if allowed to progress at a student-set pace, are examples of why it has been difficult for PSI advocates to convince administrators to implement large scale (college-wide) applications.

In Lindsley’s (1992) commentary on why effective instructional practices such as PSI have not been widely adopted, he noted that even though there is a consensus that the U.S. education system is far less than optimally effective (some would say failing; Sherman, 1992), the motivation for change is not present. He asserted this is not because America does not need educational practice reform—they clearly do—it is because they do not want it, or at least they do not want the answer research has provided for them: that effective education requires considerable effort, both on the part of the student and the instructor.

Sherman (1992) identified what he called “resistance to change” as the main obstacle to PSI’s adoption. The educational establishment, Sherman observed, does not want change. He implied there are possible financial reasons for this resistance, in terms of people in power (government and/or industry presumably) having “conflicts of interest.” He also said the role of the teacher in PSI is so very different from what people
are used to, that this also created problems. Buskist et al. (1991) also cited inertia as a main reason for the decline of PSI and made the useful point that lecture-based methods may very well be maintained as a function of intermittent reinforcement because of their seeming occasional successes. In their words, “people still learn; they become more or less educated. In spite of the lecture system, many students become doctors, lawyers, politicians, business people or educators who enjoy the respect of their communities” (p. 230).

Other obstacles to implementing PSI have been identified in the literature, including lower course completion rates, increased time on task, procrastination, and social problems resulting from a lack of interaction between teacher and student (Ainsworth, 1979; Meek, 1977). Challenges of the greater time required for initial development and a bimodal grade distribution (as opposed to the usual bell curve) are other issues instructors and administrators have identified (Boyce & Hineline, 2002; Fox, 2004). However, Kulik et al. (1979) looked for and did not find evidence for many of these assertions in their review of the PSI literature up to 1979.

Buskist et al. (1991) point out two other possible reasons for resistance to PSI implementation: time and ease of use. Many instructors believe lectures are the easiest and least time-consuming way to relay information. This pro-lecture bias may exist because instructors are most familiar with this format and using it does not require spending additional time learning a new methodology. Lecturing is also seen as a way to maintain the status of instructors as powerful and all knowing. These are hard traditions and assumptions to change.
Finally, several critics have claimed PSI could not be used to teach higher order or critical thinking skills, and is only effective for memorization of definitional subjects (Meek, 1977). Reboy and Semb (1991) addressed this issue by arguing that PSI is an instructional methodology and, as such, not responsible for course content, and by showing that students in PSI courses did in fact show improvements in higher-order thinking skills. Pear (2002) and Pear and Martin (2004) have dealt with this issue specifically by arranging PSI course material in a way that requires higher order thinking, in accordance with a revised version (Crone-Todd, Pear, & Read, 2000; Pear, Crone-Todd, Wirth, & Simister, 2001) of Bloom’s taxonomy (Bloom, Englehart, Furst, Hill, & Krathwahle, 1956).

To address some of these implementation issues and criticisms, many studies have looked at the effects of modifying the PSI approach. Some such arrangements advocate the use of certain components of PSI in a more flexible format or with a different mode of delivery. The next section will look at some of these possibilities.

**Modifications of PSI**

PSI has been modified in many different ways to suit differing educational needs and to help facilitate its efficiency, usually as measured by reduction of cost and time (Hursh, 1976). These different approaches fall into two main categories: modification of component use and computerization. Component use refers to the use of one or several components of PSI, but less than all five as originally outlined by Keller. Computerization refers to the use of a computer for the presentation of material or quizzes, the presentation of feedback, and/or computerized grading and record keeping of
student progress. Computerized forms of PSI may or may not modify component use. Many computer-based PSI programs utilize all components of Keller's plan (e.g., Pear & Novack, 1996). In other words, the only modification to the system may be the addition of the computer itself. First, a review of component modifications is necessary to look at the various arrangements and the research on their efficacy. Then computer-based modifications will be examined in the same way.

Component Use Modifications

There are several distinct yet connected components of PSI. Keller (1968) originally identified five key components: mastery requirement, self-paced progression, printed study guides that direct student learning, use of student proctors, and occasional lectures for motivational purposes. Subsequent reviews and studies continued to describe PSI's critical features in much the same way (Kulik et al., 1979; Robin, 1976), but more recent analyses have either focused on different aspects of the components or have described them differently. For example, Worland (1998) also identified five key components, but instead of the occasional lecture component, he described frequent testing on small units of material as a distinguishing factor. Frequent testing is described as part of the mastery component in most earlier studies, as is clear in the Robin (1976) and Kulik et al. (1979) component use reviews. However, it may be useful to consider PSI's frequent testing procedures independent of the mastery criterion for reasons such as the "testing effect," which will be discussed in more detail in the next section.

Fox (2004), in one of the most recent assessments of the salient features of PSI, advocated a more progressive component breakdown, with a focus on the most likely and
advantageous applications. He did this in light of our current understanding of which components have been shown most necessary and how current technology may affect PSI’s implementation. Fox described these essential elements as unit mastery, flexible pacing, on-demand course content, immediate feedback, and peer tutoring.

Unit mastery refers to the requirement that students must demonstrate mastery for each unit before moving on to the next unit in the course. The time it takes to successfully master a unit varies across students, necessitating flexible pacing within the course. In traditional PSI, the pacing is entirely left up to the learner, but in subsequent modifications, several strategies have been implemented to in some ways restrict the time allowed to complete the units in a PSI course, primarily for the purpose of discouraging procrastination. This being the case, Fox identifies the need to maintain some level of individualized control such that at least a faster progression through the course is possible.

On-demand course content refers to a slight alteration of the original “focus on the written word.” On-demand reflects the ability of students to access content materials at any time, particularly via technological means such as web-based course management systems. Immediate and detailed feedback has been a consistent component of PSI since the beginning (Kulik et al., 1990) and continues to prove necessary for its success (Fox, 2004). The difference between early applications and more recent ones lies in how the feedback is provided. Traditionally, feedback was given by a proctor after manually grading the student’s quiz. Today both the grading and feedback can be more efficiently accomplished either by or via a computer. The addition of Internet capabilities also allows these functions to be performed off site, from literally anywhere an Internet connection is
available. More of the advantages of technological advancements in instruction will be discussed in subsequent sections.

As a result of the evolution of PSI's component arrangement, it is now more clear which aspects are most essential and how they might be best implemented. The component analyses reviewed outline the need for frequent and direct assessment of mastery; some type of flexible pacing within the course; thorough, well-designed, and easily accessible content; and the need for immediate and detailed feedback. Proctoring has been shown to be a less necessary component of PSI (Caldwell, 1985), and tutoring, while potentially useful, has not been demonstrated as essential. Each of these components is important, but not equally so. Mastery is not only a component, but also the ultimate goal of PSI. Pacing, on-demand content, and feedback are only important to the extent that they contribute to the aim of student mastery. As such, the next section will outline some of the most important features of mastery.

Mastery. The component analysis literature cited previously and several other reviews (Buskist et al., 1991; Hursh, 1976; Kulik et al., 1978; Sherman, 1992) conclude that there is an abundance of evidence for the effectiveness of the mastery component in PSI. The definition of mastery, however, is not as clear. There are three variable points: (a) how mastery is measured, (b) the criterion level that identifies when mastery is reached, and (c) the practice of instructor-controlled versus learner-controlled progression rate.

The measurement of mastery has been traditionally assessed by the number of correct responses on a written test. Some suggest, however, that rate or fluency of
responses be added to that definition (Binder, 1996). For example, such a definition would require both a minimum percentage of correct responses (e.g., 90% correct), as well as a time limit within which to complete either each question or the whole test (e.g., 1 minute per question or 10 minutes per quiz).

The effects of adding a fluency requirement in some learning situations has been shown to be positive (Binder, 1996), but the research on its utility in mastery learning has shown mixed results. One study by Luyben, Hipworth, and Pappas (2003) reported improved academic performance and attitudes when using elements of both mastery learning and fluency training over mastery alone. In contrast, two other studies by Wong (1999) and Covington (2006) found no additional benefits to adding a fluency component to mastery. As such, it is currently unclear that adding a fluency requirement to mastery learning is of necessary value.

The criterion level at which mastery should be measured is another aspect that has seen considerable variation in the literature: PSI’s original format used a 100% correct criterion most often, but as the system was modified and applied to different situations, the criterion shifted from an average low of 80% up through the original 100% (Kulik et al., 1990). Kulik et al. (1990) stated in their review of mastery programs that studies requiring high levels of mastery showed higher effect sizes than those with lower criterion requirements. Based on Kulik et al.’s categorizations, it would seem that a criterion of at least 90% is advisable.

Finally, the use of instructor-paced progression rates in a course sets up a very different structure than one in which the learner progresses at their own self-paced rate. Traditional PSI was completely student paced, meaning the student progressed through
the course as he or she completed the unit quizzes, with no restrictions on time or allotted number of quiz attempts. However, due to concerns of student procrastination and subsequently lower course completion rates, much of the modified PSI research contains examples of courses with instructor-imposed limitations, either on the number of assessments allowed per unit, or on the time frame allotted to take the assessments.

Kulik et al. (1978) identified eight studies that compared the use of teacher-defined deadlines with courses that did not have deadlines. They found that setting deadlines did not have a significant effect on outcome measures and that limitations on self-pacing did, in fact, decrease procrastination and withdraw rates. It seems placing restrictions such as these on the student's progress does not hinder the effectiveness of PSI interventions, giving reason to believe that mastery can still be achieved without complete self-pacing.

However mastery programs are implemented, the underlying function of mastery is to create a condition under which student performance improves as a result of attempting to meet a minimum criterion of expectation. Johnston and O'Neill (1973) demonstrated the power of setting a minimum criterion for performance, as mastery does, on student achievement. Five experiments were carried out to show the varying effects of criteria expectations and associated grade label on student performance. In all five experiments, students in an undergraduate, junior-level abnormal psychology course attended individual performance sessions in which they met with a performance manager. During these sessions, the students were given cards with fill-in-the-blank questions on them related to the unit on which they were currently working. Students agreed to a time frame for answering questions, usually about 10 minutes, and the performance manager
kept track of how many questions were answered correctly and incorrectly during this
timed period. A rate of correct and incorrect answering performance was calculated and
graphed at the end of each session.

In the first experiment, no criteria for performance were specified. The students
were told to "do their best" and that they would be graded "on a curve" at the end of the
semester. The result was consistently low levels of student performance. The averages for
this condition were 1.3 incorrect responses per minute and 3 correct responses per
minute.

In the next three experiments, performance criteria were specified and designated
for coding and reporting as high, medium, and low levels of expectation. The "high"
criteria were 3.8 items per minute correct and 0.4 items per minute incorrect, the
"medium" criteria were 3.1 correct and 1.1 incorrect, and the "low" criteria were 2.5
correct and 1.7 incorrect. The criteria conditions in effect at any one time were varied for
each student, but all conditions, provided they were met, specified a grade of "A." The
results from these experiments showed again that student performance very closely
matched the expectations of the criteria level set and also showed that students performed
better when criteria were clearly defined (as compared to the first experiment).

In the fifth experiment, the high, medium, and low levels were again specified, but
each level corresponded to a different grade (A for achieving high criteria levels, B for
medium, C for low). Again the results demonstrated that the expectation of performance
specified by the criteria that were set controlled behavior to a very high degree. In fact,
students would change their behavior to meet newly specified criteria, even if it meant
decreasing the quality of performance previously attained or currently being attained (e.g.,
a student currently answering 4 questions correct per minute would change his behavior
to answering 2 questions correct per minute if the criteria for an A was changed from 4
correct per minute to 2).

Caldwell (1978) showed results similar to those of Johnston and O’Neill (1973),
stating that “the results of these studies strongly suggest that students will generally
perform at the level required of them” (p. 65) and concluding that “of the five essentials
[referring to PSI’s five components], mastery is the essential essential” (p. 65). In
addition to the clear conceptual benefits of mastery, there may also be benefits to the
procedural format. The mastery learning format allows for and often involves multiple
testing attempts. This multiple-retake function is a way to facilitate the different skill
levels and needs of the students attempting to master the material by allowing additional
practice/study for those that need it, without requiring it of those who do not. In addition
to this technical role, multiple testing opportunities may provide a benefit in and of
themselves. The following section will focus on this research.

Frequent testing. In addition to the positive effects of frequent testing found in the
mastery learning literature (Hursh, 1976; Kulik et al., 1979; Robin, 1976), research on the
“testing effect” (Bangert-Drowns, Kulik, & Kulik, 1991; Glover, 1989; Leeming, 2002;
McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006) lends further
support to the value of providing multiple testing opportunities. These studies show, in
general, that increasing the frequency of testing produces a positive effect on both short-
and long-term retention compared to conditions that include studying behavior (such as
reading) without testing.
A recent and representative investigation of the testing effect was conducted by Roediger and Karpicke (2006), in which two experiments showed large effects of testing on retention. Experiment 1 demonstrated that retention improves when subjects are given a test immediately after studying as compared with a group who studied the same material without being given a test. Retention was measured with a free-recall test 2 days after exposure and again 1 week after exposure. The results after 2 days were 68% correct recall for the group receiving the test versus 54% for the non-tested group. After 1 week, the tested group recalled 56% of the material correctly and the non-tested group recalled 42%.

In experiment 2, three groups were compared. Group one was instructed to read a passage, four times in 5-minute intervals with 2-minute breaks between intervals; this group was referred to as the SSSS group, to indicate the presence of the studying intervention in all four intervals. Group two was instructed the same way, but studied for only three intervals, followed by a test in the last interval; this group was referred to as the SSST group, indicating the presence of the studying intervention in the first three intervals and a test in the fourth. The third group was instructed to read the passage during the first interval and then given three free recall tests, also in 5-minute intervals with 2-minute breaks between. This group was referred to as the STTT group. During the breaks, subjects were instructed to do multiplication problems as a form of distraction to eliminate the use of rehearsal for memorization. All groups were given a final test (also free recall) either 5 minutes after the last interval or 1 week later. The results for long-term retention (1 week later) were as follows: for group one (SSSS), 40% of the passage was accurately recalled; group 2 (SSST), 56%; and group 3 (STTT), 61%. The difference
between group 2 and group 3 was marginal, but the difference between groups 1 and 2 was significant, \( t(58) = 3.21, d = 0.82 \), as was the difference between groups 1 and 3, \( t(58) = 4.78, d = 1.26 \).

In another analysis of these data, Roediger and Karpicke (2006) looked at absolute measures of forgetting, which were calculated as \( \frac{\text{initial recall} - \text{final recall}}{\text{initial recall}} \). This analysis found an even starker contrast between groups. Group one (SSSS) forgot 52% of the passage they had read a week earlier. Group two (SSST) forgot 28% of the passage, and group three (STTT) forgot only 14%. While these results are impressive, it is unclear to what they are attributable. The authors refer to several theoretical explanations such as transfer appropriate processing, elaboration of existing memory traces and their cue-target relationships, and the introduction of “desirable difficulties.” These theories are not elaborated on with any behavioral or measurable distinctions and as such are not very helpful in understanding the processes resulting in the testing effect. It seems justified to say there is a relationship between writing about material we have read in a testing situation and our long-term retention of that material (long-term being defined as a week in this instance). The parameters under which these effects hold true would require further testing, which the authors indicate as well. Despite the ambiguity surrounding the specific explanations of the testing effect, the results are still useful in that they provide another example of the benefits of an instructional system that utilizes frequent testing.

In a review by McDaniel et al. (2007), three experiments were analyzed that attempted to generalize the applicability of previous findings on the testing effect. Most experiments of the testing effect used contrived word lists as learning materials. It was
thought that by using more educationally relevant materials found in most college courses, such as lectures and articles, that the results would also be more relevant. The findings were consistent with past research in that all results demonstrated the testing effect, regardless of material or environment (laboratory vs. classroom). They also found that short-answer tests produced a greater effect than multiple-choice tests on final comprehension and retention scores and that immediate feedback also had a positive effect.

Most testing effect research has measured retention using verbal and written tasks, but some research has shown that retention of visuo-spatial information can also be enhanced through multiple testing opportunities. In a study by Carpenter and Pashler (2007), two groups were directed to study pictorial maps, one through conventional study and one through computer-prompted tests. The repeated test group was able to identify and draw in missing features significantly better than the group that studied the map without testing. The results seem to suggest that the testing effect is not limited to discrete verbal tasks, which has implications for a more diverse field of application.

The conclusion that can be drawn from both the testing effect literature and the extensive PSI modification literature is that mastery and its provision for frequent assessment is an integral component in effective instruction that should arguably be at the forefront of any educational intervention. The implementation and delivery of any such intervention is another important consideration. The use of the computer and later the Internet to improve the usability and functionality of mastery learning programs is the most recent approach and will be examined next.
Computer-Based Modifications

The addition of computer technology to the mastery model can be viewed as part of the progression toward an attempt to make the model more effective and efficient. Sherman, in his 1992 review and commentary, asserted that PSI and computer-based instruction (CBI) have much in common. The main difference, Sherman said, was that CBI was much more widely accepted because it is more likely to be viewed as a supplement to "traditional" instruction, a much less threatening proposal than the replacement of it.

CBI itself has been shown to be an effective educational strategy (Desrochers & Gentry, 2004; J. A. Kulik, 1994; Kulik & Kulik, 1991; Schacter & Fagnano, 1999). Desrochers and Gentry (2004) reviewed many CBI applications and found CBI to be an effective educational approach for both assessment of individual student needs and flexible delivery of content. Kulik and Kulik (1991) also looked at the effectiveness of computer-based instruction in a meta-analysis and reported overall greater effect sizes for computer-based instruction compared to traditional lecture-based instruction. CBI, in combination with mastery learning, has also demonstrated positive effects (Montazemi & Wang, 1995; Pear & Novak, 1996; Pear & Martin, 2004; Skinner, 1990).

A current example of the merger of mastery learning and CBI is the Computer Aided Personalized System of Instruction (CAPSI), developed by Joseph Pear at the University of Manitoba (Pear & Novak, 1996). CAPSI is a relatively traditional implementation of PSI with the aid of a computer for the presentation of quizzes and many administrative functions. Pear and Martin (2004) stated that all CAPSI courses
present specific study objectives, provide contingencies for verbal activity related to the objectives, monitor student progress, and provide feedback. The computer program does this by sending students essay-question quizzes when requested via email and then sending the completed quiz to two peer reviewers (students currently taking the course who have mastered the quiz), or, if peer reviewers are unavailable, a TA or the instructor will grade the quiz within 24 hours. The quiz is then sent back to the student with a "mastery" or "restudy" evaluation as well as commentary on both correct and incorrect answers.

CAPSI necessitates the use of proctors or TAs to grade and give feedback on the quizzes because of the essay nature of the questions. Pear defends this additional administrative effort and cost by asserting that the benefits of higher-order thinking skills the students gain outweigh such costs (Pear, 2002), as have others (Conner-Greene, 2000). In another example, Skinner (1990) demonstrated the effectiveness of CBI in a PSI course. He found that low-achieving students benefited more than high-achieving students, but that all benefited from CBI over non-CBI conditions.

Brothen and Wambach (2000, 2001, 2004) performed a series of studies that looked at the effectiveness of computer-based quizzes in a PSI introductory psychology course. These studies examined differences in effectiveness of student use strategies and the effects of time limits on quizzes. The authors contended, as did Pear with CAPSI, that the proctor function should be maintained, but can be made more effective with the addition of the computer for test administration functions. Brothen and Wambach pointed out that this added efficiency frees the proctor or instructor to focus more on interacting
with the students through feedback and “to work more comprehensively with students to help them develop as effective learners” (p. 255).

In the first study, Brothen and Wambach (2000) had students take chapter pretests that had to be mastered at 90% before taking chapter quizzes. The chapter quizzes could be taken up to five times, the highest score counting for the final chapter test grade. The results showed that students’ pretest scores improved significantly as they continued to make mastery attempts and that students taking fewer chapter quizzes earned significantly higher quiz scores across all chapters.

In their second study, Brothen and Wombach (2001) looked at the different ways students used chapter quizzes, either as a means to gain feedback on level of individual preparedness or as a means to study and how these approaches affected exam scores. It was hypothesized that a “prepare – gather feedback – restudy” strategy for taking quizzes would lead to higher unit exam scores than a “quiz to learn” strategy, whereby students repeatedly take quizzes, in lieu of reading or studying, in an attempt to learn the material for the exam. In this experiment, students took chapter quizzes as many times as they liked and their highest score was recorded. No specific mastery requirement was enforced for progression to the next unit, but the grading scale was such that 8 out of 10 multiple choice questions had to be correct for any points to be earned. As such, a mastery criterion of 80% was technically in effect. Four unit exams were given every five to six chapters throughout the course. There was a significant correlation found between number of quiz attempts and quiz score. The greater the number of quiz attempts made on each unit, the lower the final quiz score. There was not, however, a significant correlation between number of quiz attempts and exam score. The average correlation between all

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
students' mean scores on quizzes taken prior to each unit exam and their respective unit exam scores was significant, $r(29) = .39, p < .05$. The mean score on all quizzes taken by students correlated with their total points on unit exams, $r(29) = .53, p < .01$, and course grade, $r(29) = .62, p < .01$.

In a third study, Brothen and Wombach (2004) examined the effects of setting time limits on the quizzes, again to encourage adequate preparation before making a quiz attempt. The results were significantly higher exams scores for students who had a 15-minute time limit on quizzes as compared to those who had no time limit, $t(44) = 2.44, p < .05$.

The results from these experiments suggest that providing a structure to encourage adequate student preparation before making mastery quiz attempts may be beneficial to overall course mastery. Setting time limits on mastery attempts may be an effective strategy for promoting student preparation activities such as reading and note taking. Supporting effective student preparation activities, in turn, supports mastery.

The use of feedback, and the variations available with the use of computers in mastery learning, is another area of interest in this literature. Topics include the timing of feedback (Buzhardt & Semb, 2002); the type of feedback given (simple vs. elaborate; Chase, 2006); type of information included in feedback (praise, encouragement, comparative group information; Worland, 1998); and accuracy of feedback (Martin, Pear, & Martin, 2002). These feedback variations have interesting implications for effective instruction, but most of this research has found that basic feedback is similarly effective to elaborate feedback (giving the learner additional information, such as indicating the text to study or giving the learner additional praise or encouragement) for most
applications and much less time consuming and effortful for instructors. Basic feedback usually consists of providing the student with the number of questions correct, the correct answer for each question, and the student's answer for each question (showing whether it is correct or incorrect). Basic feedback can also be automated in most course management systems, thereby making it a seemingly more logical choice for most mastery applications.

Many attempts have been made to better understand and improve upon PSI and the mastery model through component analyses and modifications of its use. The addition of the computer to implement the mastery framework and many of its administrative functions has also made it a more viable option for a wider range of applications (Pear & Martin, 2004). The potential for further benefits, especially in light of new technological advances, warrants renewed consideration for use of the mastery model in higher education. At this time, however, implementation of a full PSI framework is often not a viable option, given the pervasive and entrenched semester/term structure at almost all colleges and universities and the administrative reluctance toward widespread change. It has been established that mastery is the most integral part of PSI and that the other components, except for flexible pacing, can be quite easily integrated into otherwise traditional courses. This being the case, perhaps there is another way to achieve mastery or even a higher level of mastery than is currently being realized, within the confines of the current semester/term arrangement. If so, use of the most current technology will probably be helpful, if not necessary, and is the subject this discussion turns to next.
Web-Based Technology in College-Level Instruction

The utility of computers in the implementation of mastery programs such as PSI has been clearly demonstrated. One limiting factor, however, is the requirement that the student be in proximity to a specific computer. If specific software is required, then the student must still interact with that program at the location where the software is installed. This proximity limitation is likely to be inconvenient for both students and the support staff required to maintain a computer lab. This arrangement is especially impractical for many non-traditional students, who commute to campus a limited number of times per semester. Distance education courses would be particularly difficult to implement if students were required to use an on-campus computer lab for their coursework. One answer to this dilemma is the use of Internet or online communication.

The most common form of Internet-based educational interface used by colleges and universities is the course management system (CMS). CMSs allow students to interact with instructional materials, the instructor, and other students via any computer that has Internet access. CMSs have given rise to a multitude of opportunities for implementation of new instructional ideas (Eskicioglu & Kopec, 2002). The use of CMSs has greatly increased in the last 10 years and projections of their future use suggest the trend will continue. Many colleges and universities have purchased these systems, both for “traditional” and distance education use (Bradford et al., 2007).

CMSs are typically characterized as interactive online interfaces that are available for the student to access at any time. Each course interface, referred to often as a class website, is designed by the instructor such that everything available to the student relates
to that particular class. Some of the common features are access to course information such as syllabi, class schedule, and instructor information; access to course documents such as PowerPoint presentations and study objectives; and discussion board functions. There is also an assessment function that allows the instructor to post feedback and grades. Some additional general benefits of CMSs are as follows: students do not have to be in the same place at the same time (Pear, 2002); there is often a marked reduction in printing costs; grading and recording are automatic; students have easy access to course information, assignments, and grades; and students receive immediate feedback on performance (Barnett, 2004).

According to a report by members of the United University Professions Technology Issues Committee (Bradford et al., 2007), the potential benefits of a CMS are (a) increased availability, (b) quick feedback, (c) improved communication, (d) tracking, and (e) skill building. Availability refers to a student’s ability to view material, interact with other students and the instructor, take tests and quizzes, and view grades at any time, anywhere they have Internet access. Improved communication and tracking are more managerial aspects, but the components of quick feedback and skill building are important features with regard to compatibility with the mastery approach.

With regard to feedback, when tests or quizzes are assigned within a CMS, students can receive immediate feedback, either after each question or at the end of the test or quiz. This feedback can be automated if multiple choice, true/false, or specific short answer formats are used. The student will see which questions they answered correctly and incorrectly and any additional information the instructor adds. If essay questions are used, feedback can take on the form of a sample correct answer that the
student can view to judge similarity and likely congruence with their answer. The test would then be graded by the instructor, TA, peer reviewer, etc., and posted back to the site for the student to view. Skill building refers to the ability to present activities, quizzes, and other review materials in a format that is easy for the student to access and the instructor to deliver multiple times in an individualized manner. This is especially important with regard to mastery, as frequent and repeated access to instructional materials and assessment functions are aspects that have been difficult to deliver and implement using the traditional classroom approach.

The most popular commercial versions of CMSs used today are Blackboard Learning Systems and Web Course Tools (WebCT), which due to a recent acquisition are now owned by the same company. In a press release from Market Data Research company (MDR), highlights from their annual college technology review for the 2005-2006 school year reported that most institutions (92%) utilize some type of CMS, with the most popular being Blackboard (used by 47% of colleges) and WebCT (30%). Since Blackboard and WebCT are now one, Blackboard is used by nearly 80% of all colleges using a CMS.

The 2006 National Survey of Information Technology in Higher Education states that campus IT officers report the single most important IT issue for their institutions over the next 2-3 years is network and data security, as it has been for the last 3 years (2004-2006). This is potentially important to this discussion because in the 4 years prior to this (2000-2003), instructional integration of IT was the #1 issue (Green, 2006). This may imply that the people in charge of IT at colleges and universities are not necessarily focused on or working toward the use of IT in instruction. This being the case, it may
often be left primarily up to instructors to decide how IT is used for instructional purposes.

The use of online technology for education in general is fairly new (Aggarwal & Legon, 2006). With regard to mastery applications, it is just beginning to be explored, but there are a few studies that have attempted this union. Koen (2001) summarized what he asserted was the first web-based PSI course implementation. He began using the online PSI in 2000 for a freshman engineering computer class and reported 16 of 20 students finished the course on time and 3 of the remaining 4 finished the following semester. In 2001 he reported an 80% completion rate for 60 students taking the course during Spring semester. It is not clear exactly what kind of online platform was used in the creation of this course, but the author describes its attributes in ways that are consistent with the previous discussion of CMSs, albeit a bit more customized (references to live video chats and frequent use of instructor-generated instructional video segments).

Eppler and Ironsmith (2004) implemented a web-based mastery course for their Human Development classes. They compared web-based PSI to traditional lecture for 814 students over four semesters. The mastery criterion was set at 90% and they allowed only four mastery quiz attempts per unit. To reduce procrastination, they imposed deadlines on the completion of each unit with strict consequences (loss of a letter grade for each unit not passed on time). The authors then allowed for one missed deadline to be “bought back” at the end of the semester with the accumulation of bonus points for activities such as early unit completion, passing a mastery quiz on the first try, mastering units at 100%, attending review sessions, and completing practice exercises. web-based PSI produced significantly higher average course grades and final exam scores each semester, yielding a
large average effect size of $d = .87$. Course withdrawal and failure rates were very similar between groups (20% for lecture, 26% for web-based PSI).

Citing academic integrity as a rationale, Eppler and Ironsmith (2004) chose to require students to come to campus to take the unit mastery quizzes under the supervision of a proctor. This practice is consistent with traditional PSI implementations, but discounts an important advantage of online presentation when used for distance education: the advantage of being able to offer students the ability to fully participate in the course remotely. More traditional, on-campus courses could also benefit from the ability to use self-monitored quizzes, as proctoring requires additional resources often unavailable to many instructors. The authors do not address this point directly, but suggest in the discussion that future online courses could be modified to allow for greater off-campus participation.

Brinkman (2007) is the most recent publication of an attempt to combine web-based technology with a mastery-based pedagogy and the only currently published study expressly using a CMS for delivery. This study compared two different methodologies (PSI vs. lecture) and modalities (CMS vs. classroom) for college instruction. A course in discrete math and statistics was taught in two parts. The discrete math portion was taught using a PSI method and delivered through a CMS (WebCT). The statistics portion was then taught using a traditional lecture approach. Data were taken for four consecutive terms over the course of two academic years. Many measures were used such as online student surveys, interviews, diary studies, observations, a controlled usability test, tracking of online behavior, and academic assessment results. Many of these measures were obtained to look at student behavior differences, learning strategies and student
attitudes toward CMS use. Mean exam scores revealed a significant difference between the two parts of the course. The CMS PSI condition yielded a 14-point higher mean exam score than the lecture-taught portion for the first year and a 20-point higher mean score the second year. This difference was found to be significant using a repeated-measures ANOVA, $F(1, 332) = 424.09, p < .001$.

One problem with drawing conclusions about the effectiveness of mastery use with CMSs in this study is that the mastery criterion was rather loosely defined. The mastery quizzes were 5-question multiple-choice quizzes that had to be attempted for access to the next module quiz, but not completed to a specific criterion of mastery. The students were told to demonstrate mastery through use of the self-tests before requesting the next module, but it was not formally required by any actual demonstration of mastery (Brinkman, 2007).

It was stated in the student interviews that many students did in fact take the self-tests to mastery at a self-imposed criterion of 100%. Many students even stated that they found this “requirement” to be unreasonably stringent. This feedback indicates that some students did, in fact, use the self-tests as demonstrators of mastery, even though it was not required (Brinkman, 2007). This may be what accounted for the difference between the exam scores of the two groups. While this is an interesting possibility, it is difficult to pinpoint the exact variables responsible for the difference in exam scores, and indicates a need to identify those variables and actually test them in a more methodologically rigorous way.

Effective instructional practices, like mastery learning, and potentially more effective modes of delivery, like course management systems, present an enticing
combination for new educational strategy ideas. While the most effective means of using such a combination is yet to be clearly demonstrated, another essential component is a willing instructor. This is not something we should take for granted. For example, PSI has been shown to be an unequivocally superior instructional method (to lecture) since the 1960s, and yet its use has not only failed to supplant the lecture, it has substantially declined (Sherman, 1992). As has been discussed, there are many potential reasons for this, but none more important than a lack of focus on instructor needs.

Instructor Adoption and Implementation

Instructor implementation is an important issue to consider when advocating a new educational program or intervention. In the case of mastery learning, setting up even a modified PSI course can be effortful and time consuming (Boyce & Hineline, 2002). It can be difficult to convince instructors of the utility of spending extra time to learn a new program (especially associated with computer use) with the assertion that it will save them time “down the road” or to reorganize a course that they feel is already meeting their needs. Some instructors are also reluctant to venture into the realm of “new” technology for fear the learning curve is just too steep (Bongalos, Bulaon, Celedonio, de Guzman, & Ogarte, 2006). Another reluctance may be due to doubt about the impact of new technology on actual student outcomes (Kulik et al., 1980). These are real and valid concerns that must be addressed if the application of empirically valid educational research is to be a realistic goal.
In a study by Nijhuis and Collis (2005), many instructor concerns with changing expectations are identified as inherent in the changing face of higher education. The following statement summarizes this connection.

As universities transform into enterprises, academics are facing new challenges, especially in their teaching. This is because of the demands for student-centered programs that offer more flexibility, the use of Course Management Systems such as Blackboard, and the expectation that instructors will perform (more) efficiently and effectively. (p. 1)

The authors of this study suggested support for instructors in the form of human performance technology (HPT) interventions. HPT in this instance involved using a task management system to help instructors manage the increased demands being placed upon them by the college’s expectations to keep up with new technology, greater accountability due to budget constraints, and changing demands by students.

Another study by Bongalos et al. (2006) reported on a group of teachers attempting to implement a CMS into their teaching and course structure. This is an interesting study as it addressed many of the implementation issues discussed so far, but also examined the impact of those issues on college instructors in a developing country (the Philippines) where access to technology is of special concern and importance (due in part to greater geographical distance between prospective students and universities). They found the instructors had many implementation difficulties with regard to learning and becoming comfortable with the new technology. However, they also found they were open to and excited about the possibilities CMSs have to offer and were willing to learn. One comment was that the “courseware (was) helpful and user-friendly, but . . . it should never replace the teacher” (p. 702). The authors’ conclusion was summarized well by the following statement:
No innovation has a realistic chance of succeeding unless teachers are able to express, define and address problems as they see them, unless teachers come to see the innovation and change as theirs. The ultimate outcome of the innovation... depends on when and how teachers become part of the decision to initiate them. (Bongalos et al., 2006, p. 1)

There are many issues related to instructor interest in, and time to, initiate new methodologies. In a study by Jones and Jones (2005), perceived effectiveness of Blackboard was measured for both students and faculty. They reported that all responses but one, by both students and faculty, were significantly positive.

The findings of this implementation literature are important because they help answer social validity questions that are sometimes ignored in instructional research (Wolf, 1978). Other studies have looked at perceived usefulness, ease of use, and acceptability of information technology (Davis, 1989); educational achievement and social concerns (Johnson, 2005); and implementation of new instructional approaches in general (Aggarwal & Legon, 2006; Eskicioglu & Kopec, 2003). Davis (1989) found a strong correlation ($r = .63$) between perceived usefulness and actual use behavior, stronger even than the correlation between perceived ease of use and actual use ($r = .45$). These results suggest a need to focus on utility when making a case to instructors about why they should invest the time and energy necessary to learn a new instructional strategy.

Along with instructor considerations, there are also issues of time, money, and effort expended by the administration to introduce and maintain technology. Technology is an indispensable part of all colleges and universities today and, as such, administrators want and need instructors to take advantage of the technology as the institution invests in
it. A lot of money is spent keeping a college or university technologically current. If instructors are not using this technology or not using it to its full potential, the institution's return on investment will be diminished (Pillman, 1990; Yohon, Zimmerman, & Keeler, 2004).

Rationale for Current Study

Research supports mastery learning, the modification of mastery programs like PSI to take full advantage of the most valuable components (mastery criteria, frequent testing, availability of content, and immediate feedback), and the use of technology in the college classroom. However, there has been very little research on the combination of these strategies (Pear, 2004; Brinkman, 2007; Eppler & Ironsmith, 2004). And even fewer studies have addressed implementation issues (Brinkman, 2007).

The question is not whether technology is going to be a part of higher education instruction; it is whether sound educational methodology is going to be at the base of it. This study proposes a format that not only utilizes both technology and proven instructional techniques, but also one that can be easily and successfully implemented by almost any college instructor.

The current study evaluated a format that uses a course management system, already in place at most colleges and universities, to encourage mastery of course content through frequent, online chapter review quizzes. This structure was designed to increase the likelihood of mastery by making grades contingent upon successful completion of the review quizzes.
It was the intention of this study to make the addition of a mastery condition to an otherwise traditional course as simple and straightforward as possible. The effectiveness of this strategy was measured by chapter and final exam scores as compared to control conditions, and social validity data was taken to examine student attitudes toward this "new" approach.

METHOD

Participants and Setting

Thirty-two students enrolled in two sections of an introductory psychology course at a midsize community college in the Midwest served as participants for this study. Introductory classes were chosen because they are representative of average first- and second-year college classes, which should serve to increase the generalizability of the findings. Demographic data taken by survey revealed no apparent differences between sections. A synopsis of the results follows. Students in section 1 had a mean reported GPA of 3.4, and students in section 2 reported a mean GPA of 3.1. The racial breakdown for the two sections was comparable (24 white, 2 African American for section 1, and 23 white, 2 African American for section 2). The current enrollment breakdown was balanced (mean of 12.5 credits for section 1, and 13 credits for section 2). Current class level was also balanced (20 freshmen and 5 sophomores in section 1, and 22 freshmen, 4 sophomores in section 2). Comfort level with the computer and Internet were very similar between sections (19 were very comfortable, 9 were somewhat comfortable, and 1 reported being somewhat uncomfortable in section 1, while section 2 yielded 17 very
comfortable, 9 somewhat comfortable, and 2 somewhat uncomfortable responses). Finally, in section 1, 17 students reported being traditional students and 11 reported being non-traditional, while in section 2, 22 reported being traditional and 5 non-traditional.

Section 1 consisted of 32 students, of which 15 met the criteria for inclusion at the end of the semester, and section 2 had a total of 33 students, of which 17 met the criteria for inclusion in the study. The inclusion criteria were that the students remain enrolled for the entire class, that all examinations were taken and that consent was given to have their data used, by signing a consent document (see Appendix B).

The classes met three times a week (Monday, Wednesday, Friday) for 1 hour each. All classes were held in standard college classrooms equipped with a computer, a projector, and a seating capacity for approximately 40 people.

Materials

Textbook

The textbook used for the Introduction to Psychology course was *Invitation to Psychology, 4th edition* (Wade & Tavris, 2008). PowerPoint presentations were used to cover material related to the textbook chapters during lecture. The PowerPoint presentations for all units were available on the course website for the entire duration of the course.
Course Website

A course website was created for each of the class sections using the Blackboard course management system. Students had continuous and unlimited access to their section’s site for the entire semester. The Blackboard site included the following content areas, accessible via buttons of the same name: (a) Course Information, including the syllabus and instructor contact information; (b) Lectures, including the unit PowerPoint slides; and, (c) Assignments, including the mastery quizzes (described below). The students also had access to their grades at any time via Blackboard’s grade book.

Mastery Quizzes

Four or five mastery quizzes (depending on which section the student was in), consisting of 10 multiple-choice questions, were available for half of the nine total units covered during the semester. The quizzes were accessed only via Blackboard and were available only for the duration of the unit currently being covered in class. The quizzes included questions pulled randomly from a test bank of approximately 100 questions per unit. The mastery quizzes were referred to as “review quizzes” in the syllabus, on the Blackboard site and in class. The mastery quizzes were worth 10 points each for a total of 40 or 50 possible points (depending on the section), which is approximately 8% of the total grade.
Unit Exams

The unit examinations consisted of approximately 10 essay questions each and were given for each of the 9 units (see Appendix C for a sample unit exam). The unit exams covered material in the textbook and from lecture. Unit exams were worth 40 points each, for a total of 360 points, which is approximately 66% of the total course grade.

Final Exam

A cumulative final exam was given during the last week of the course. This final exam was also in essay format and worth 40 points, approximately 7% of the total course grade.

Additional Points

Additional point opportunities came from reading quizzes, which consisted of one short-answer question about some main aspect of each new unit being started. The reading quiz was handed out at the beginning of class, before any discussion or lecture. Reading quizzes were only given the first day a new unit was started and were worth 5 points each, for a total of 45 points, 8% of the total grade (see Appendix D for a sample reading quiz).

Finally, there were three assignments given during the course: The first was a two-part assignment given the first week of class. The first part was a one-page paper answering the question “How do I learn?” The second part was a study skill assignment
requiring the students to apply some of the study skills discussed during lecture in two of their current classes for 2 weeks.

The second assignment was a behavior change project in which students choose a behavior to modify using the principles of reinforcement and the three-term contingency diagram. They were instructed to observe and record the behavior for 8 days and write up their results, including a graph showing the behavior change.

The final assignment was an end-of-semester review project in which students were required to find one example in the media that illustrated one of the main concepts from each of the units covered during the semester. Each of these assignments was worth 20 points, for a total of 60 points, which is 9% of the total course grade (see Appendix E for an example of an assignment).

Survey

Demographic data (gender, ethnicity, etc.) were obtained anonymously using a survey given out in class after the drop/add period expired (see Appendix F). A social validity survey was given to all students during the penultimate week of class (see Appendix G). It asked for information about students' experiences and assessments of interacting with the mastery quizzes in the course. It asked the students to rate, on a scale from 1 to 5, how much they agreed with each of the statements. Students were also allowed to specify at the end of the survey what they would change, if anything, about the mastery quizzes.
Experimental Design

This study used an alternating treatments design (Cooper, Heron, & Heward, 1987) to compare exam performance in two conditions. The conditions were counterbalanced across two sections to account for variations of difficulty in the chapters covered throughout the semester (see Table 1 in Appendix L for the condition assignment outcome).

Procedures

The course consisted of 9 instructional units. Each unit included four class meetings, spanning 7 days. A typical unit proceeded as follows: Day 1 was devoted to lecture, Day 2 to an application activity or demonstration, Day 3 to small group work and review, and Day 4 to the unit exam (see Appendix H for course calendar).

Before the course began, each of the units for one course section were randomly designated either mastery or non-mastery, with the constraint that each condition occur for no more than two consecutive units. The second section of the course was assigned to the conditions in a counterbalanced order. For example, if Section 1 was in the mastery condition for a particular weekly unit, Section 2 would be in the non-mastery condition for that same unit. As a result, Section 1 of the course was exposed to 4 weekly units taught with mastery quizzes and 5 weekly units taught without mastery quizzes. Section 2 was exposed to the exact opposite, 5 units taught with mastery quizzes and 4 units without.
Each class was given instructions, both verbally and in the syllabus, for taking the mastery quizzes via the course website in Blackboard. Each student also received a quiz schedule, as part of the syllabus, outlining the units for which they would be responsible for taking quizzes. Students had from 8:00 a.m. the first day a new unit was introduced until 8:00 a.m. the day of the exam for that unit, to take the mastery quizzes. Students had the opportunity to take the mastery quizzes as many times as necessary to reach mastery (90%) within this time frame. No time extensions were given.

There was a time limit of 20 minutes imposed on each mastery quiz attempt. This limitation was instituted to discourage open-book attempts without prereading. Blackboard automates this function by displaying a timer on the quiz screen that shows the student how much time has elapsed since they started the quiz. When the timer reaches 20 minutes, a message is displayed stating that the time to complete the quiz has elapsed. Blackboard also tracks exactly how long it takes the student to finish the quiz and how many times it was attempted.

The setting of this limitation is in response to issues raised by Brothen and Wambach (2001, 2004) in two studies that examined differences in effectiveness of student use strategies and the effects of time limits on quizzes. Student use strategies refer to how students use the quizzes: either as a means to learn the material through repeated exposure or as a feedback tool to assess comprehension of the material. Time limits were shown to encourage increased study time before attempting the mastery quiz.

Students took the mastery quizzes on their own, at a time and location of their choice. Each time a student took a quiz, they received immediate feedback at the end of the quiz indicating the student’s answers, whether they were correct, what the correct
answer was, and the final score. This process is automated through Blackboard. It may be the case that several retakes of the mastery quiz are necessary for a student to achieve the mastery criterion of 90%. When multiple retakes are necessary, the likelihood that a student will continue with the retakes may decrease as a result of increased time and effort required to reach the mastery criterion. In an attempt to increase the motivation necessary to follow through with the retakes, a bonus of 10 points was given at the end of the semester to those who completed all quizzes to mastery. The bonus had to be manually evaluated and entered by the instructor at the end of the semester. The non-mastery condition was the same in every way as the mastery condition except for the availability of the mastery quiz. This differential availability of the mastery quiz will have the effect of making available an additional 40 or 50 points during the mastery condition that are not available during the non-mastery condition. This number of points was not likely a serious confound, as the primary dependent variable was the unit exam score, and this score is only minimally affected by the increased availability of points in the mastery condition. There is no confound when looking at the total grade per student within each section, because each student had access to the same total available points. Between sections there is a 10-point difference, but this comparison is not relevant to this design.

The student experimenter taught both sections of this course. The essay examinations were graded manually according to specified criteria in the form of an objective answer key. The grading was done by the experimenter and an independent research assistant. Intergrader reliability was assessed by both the experimenter and independent research assistants. Thirty percent of all examinations from each section were randomly selected for re-grading. Separate grading sheets were used to avoid
influence from the other grader’s assessments and all grading sheets were attached to the test (see Appendix I for an example of the grading/calculation sheet). A percentage of reliability was calculated by adding the number of agreements to the number of disagreements between graders for each question and multiplying by 100. These scores were then averaged for each test and the test scores were averaged for each chapter.

Dependent Variables

The primary measure was unit exam scores compared across conditions. Final exam scores were also examined by comparing scores for mastery and non-mastery questions. Secondary measures of mastery quiz scores, number of students who took the quizzes to mastery, and number of times quizzes were taken to reach mastery, were also collected for trend analyses.

Social Validity

No matter how effective our treatments and interventions appear, they will not be accepted, utilized, and maintained by our consumers if they are not satisfied with both the results of treatment and the treatment itself. Wolf (1978) articulated the importance of collecting social validity data by asserting that, “It seems that if we aspire to social importance, then we must develop systems that allow our consumers to provide us feedback about how our applications relate to their values and their reinforcers” (p. 213). To address this concern, an anonymous social validity survey was given to students the second to last week of class (see Appendix G).
Instructor Usability

To address the social validity of this study from an instructor’s point of view, the investigator tracked and recorded the time it took to set up the course website, create the mastery quizzes and enter the grades into Blackboard. This additional time to implement the mastery system was compared to the time required for more traditional lecture approaches and discussed in light of the impact it may have on instructor implementation choices in the discussion section.

Data Collection and Analysis

Blackboard automatically grades and organizes all mastery quiz data. Unit and final exam scores were recorded in Blackboard’s grade book by the instructor. All other grades (reading quizzes and assignments) were also posted to Blackboard as soon as they were graded. These data were exported to Excel spreadsheets and broken down into relevant categories for coding and analysis. Assessment methods included visual inspection of primary and secondary dependent variable trends as well as descriptive statistical analyses of unit exam scores across conditions and final exam scores and grades across sections.

RESULTS

Exam Scores

The mean exam scores for each course section are listed in Tables 2 and 3 in Appendix L. The mean exam scores by condition and section are displayed in Figure 1 in
Appendix L. The units for which there was a mastery requirement revealed an overall mean, out of 40 points, of 30.02 ($SD = 3.06$) for section 1, 30.37 ($SD = 2.79$) for section 2 and 30.17 ($SD = 2.76$) for both sections combined. The units for which there was no mastery requirement revealed an overall mean of 31.18 ($SD = 1.19$) for section 1, 28.95 ($SD = 2.94$) for section 2 and 29.94 ($SD = 2.49$) for both sections combined.

The individual exam scores for each student are displayed by condition in Appendix J. Visual inspection of individual exam scores by condition shows no overall difference between mastery and non-mastery conditions. The exam scores for one student, participant 20 in section 2, did reflect a clear difference between exams taken with a mastery quiz and those taken without a mastery quiz. In other words, participant 20 scored higher on all exams for which a mastery quiz was available than on the exams for which a mastery quiz was not available. This comparison is displayed in Figure 2 in Appendix L. This result, however, is unrepresentative of the majority. Two other students, participants 2 and 21, showed a possible effect, meaning all but one exam score was higher in the mastery condition than in the non-mastery condition. These graphs are shown in Figure 3 (Appendix L). The majority of results, however, revealed no clear difference between conditions by visual inspection. An example of this most typical outcome is displayed for participant 10 in Figure 4 (Appendix L).

The exam grade distribution for each section, aggregated across conditions, is displayed in Figures 5 and 6 (Appendix L). These graphs also show no differences between the mastery and non-mastery conditions.
Intergrader Reliability

The intergrader reliability score for each unit is as follows: Unit 1, 93.5%; Unit 2, 95.0%; Unit 3, 88.6%; Unit 4, 84.3%; Unit 5, 83.3%; Unit 6, 89.5%; Unit 7, 83.3%; Unit 8, 91.5% and Unit 9, 80.83%. The total reliability score for all unit tests was 87.8%.

Final Exam Scores

The mean final exam score, out of a possible 40 points, for section 1 was 28.0 (SD = 9.3) and for section 2, 30.18 (SD = 9.15), and both sections combined, 29.16 (SD = 9.02). The final examination consisted of questions from all units covered during the semester. As such, some of the questions were from units for which the students took mastery quizzes and some questions were from non-mastery units. The scores for the questions from mastery units were tabulated and compared to the scores for the questions from non-mastery units. In section 1, the mean score for questions related to chapters for which there was a mastery quiz was 16.7 out of a possible 25 points or 66.6% correct, and the mean score for questions related to chapters for which there was no mastery quiz was 10.3 out of 15 possible points or 69.4% correct. In section 2, the mean score for questions related to chapters for which there was a mastery quiz was 10.8 out of a possible 15 points or 73.6% correct, and the mean score for questions related to chapters for which there was no mastery quiz was 18.8 out of 25 possible points or 74.4% correct. For both sections combined, the average score for mastery question points was 27.4 out of 40 total possible points or 70.1% correct and the average score for non-mastery points was 29.1 or 71.9% correct.
Mastery Quiz Scores

Section 1 had a mastery quiz mean of 9.017 (SD = 1.07) and section 2 had a mastery quiz mean of 8.75 (SD = 1.36) for a combined mean of 8.91 (SD = 1.2). In section 1, 10 of 15 students completed mastery quizzes for all assigned chapters and 5 of those 10 took all quizzes to mastery, (a score of 9 or higher). In section 2, 6 of 17 students took all available mastery quizzes and 3 of those 6 took all quizzes to mastery. Students who scored 5 or lower on the mastery quizzes were very likely (99.09% of the time) to retake the quiz for a higher score. Those scoring a 6 or above, but less than 9, were somewhat less likely to retake the quiz (89% of the time).

To account for possible outcome differences for those participants who fully contacted the intervention as intended (i.e., those students who took all quizzes to mastery), a graph of averaged exam scores by unit, for students who took all quizzes to mastery in both sections combined, is included in Figure 7 (Appendix L). Visual inspection of these data reveals no discernible difference between mastery and non-mastery conditions.

Mastery Quiz Attempt

The mean number of attempts to reach mastery for each unit are listed in Table 4 (Appendix L) for both sections 1 and 2. These data show an average number of attempts between 2.1 and 6.7, but most units (6 of the 9 or 67%) had between 3 and 5 mastery quiz attempts. However, the means are not very representative, as the high standard deviations
show. Some students reached mastery in 1 or 2 attempts while others took as many as 10 or 20. Due to this variability, the medians and ranges for each unit are reported as well.

Social Validity

Results for questions 1-6 of the social validity survey for each section are presented in Figures 8-13 (Appendix L). For section 1, 9 out of 19 students agreed or strongly agreed that overall, the mastery quizzes were beneficial, while 5 out of 19 either disagreed or strongly disagreed. For section 2, 16 out of 22 students either agreed or strongly agreed and 3 out of 22 disagreed or strongly disagreed (Figure 8).

When asked whether they felt the mastery quizzes helped them learn the material faster, 7 out of 19 students either agreed or strongly agreed and 5 out of 19 disagreed or strongly disagreed in section 1. The same question for section 2 shows 14 out of 22 agreed or strongly agreed and 7 out of 22 disagreed or strongly disagreed (Figure 9).

When asked if the mastery quizzes helped them understand the material better, section 1 resulted in 9 out of 19 agreed or strongly agreed and 7 out of 19 disagreed or strongly disagreed. For section 2, 12 of 22 agreed or strongly agreed with the statement and 4 of 22 disagreed or strongly disagreed (Figure 10).

When asked if the time allowed to take the mastery quizzes was sufficient, 13 of 19 agreed or strongly agreed and 1 of 19 disagreed or strongly disagreed for section 1. The same statement for section 2 resulted in 19 of 22 agreed or strongly agreed and 1 of 22 disagreed or strongly disagreed (Figure 11).

When asked if they would like to have had the opportunity to take mastery quizzes for all units in the course, 8 out of 19 students in section 1 agreed or strongly agreed and 7
out of 19 students disagreed or strongly disagreed. In section 2, 14 of 22 students agreed or strongly agreed with this statement and 4 of 22 disagreed or strongly disagreed (Figure 12).

Finally, when asked if they would like to take another course that used mastery quizzes, 10 out of 19 students in section 1 agreed or strongly agreed and 3 of 19 disagreed or strongly disagreed. In section 2, 10 of 22 agreed or strongly agreed and 5 of 22 disagreed or strongly disagreed (Figure 13).

The final question of the survey asked, “What if anything would you change about the review quizzes?” The students were given space to write any suggestions or comments they had about the quizzes. Several interesting themes emerged. First, the students highlighted a perceived disconnect between the mastery quiz questions and the exam questions. Many comments identified student frustration with the mastery quiz questions not being the same as the study objectives for the exam. Another topic of possible significance found in the comments was that of congruence. Many students alluded to the idea that they wished the quizzes were scheduled more routinely, like for every chapter. They talk about this in terms of having difficulty remembering when to take them. These issues will be looked at more closely in the discussion section. Overall, the comments were more positive than negative. Two independent readers, the student investigator and a research assistant, rated the comments. Of the 25 comments, one rating found 11 comments were worded favorably, 8 were neutral, and 6 were negative. The second rating found 10 comments worded favorably, 9 neutral, and 6 negative, for an IOA of 94%. All comments are listed in Appendix K.
Instructor Usability Data

To assess the likelihood of instructors choosing to use the proposed system, the current investigator kept track of the time spent to implement it. Approximately 10 hours were spent setting up the Blackboard site. This included establishing the site’s structure (including the grade book) and downloading content. Approximately 20 hours were spent setting up the review quizzes in Blackboard (about 1 hour per quiz). These activities are done once for each course taught. Once this structure is established, the upkeep from semester to semester is considerably less time consuming (on average 1-2 hours at the beginning of each semester).

The current system required an additional expenditure of time for weekly maintenance. An average of 20 minutes per week was devoted to entering grades for the two classes. An average of 15 minutes per week was devoted to enabling the review quizzes. An average of 10 minutes per week was spent creating announcements and reminders for the classes. The combined weekly expenditure was 45 minutes for two classes or approximately 23 minutes per class.

DISCUSSION

The main question examined by the current study was whether the addition of an Internet-based mastery learning component to an otherwise traditional introductory college course would affect comprehension of new material as assessed on chapter exams. The results do not show a clear difference between the two treatment conditions tested. In other words, the mastery condition did not produce higher exam scores than the
non-mastery condition. While the data do not indicate a positive effect from the addition of a mastery learning component, this may have less to do with the function of mastery and more to do with how the mastery component was presented; these results may help to further clarify the important functions of instructional format and that format’s relationship to assessment.

The data on exam means show essentially no difference between the mastery condition (combined mean of 30.17) and the non-mastery condition (combined mean of 29.94). Visual inspection of individual exam scores for each student confirmed this outcome despite a few noted exceptions. The exam grade distribution also shows no significant difference between conditions. These data suggest that the added mastery component had no discernible effect on learning, either positively or negatively. This may be the case, but the amount of evidence for the effectiveness of mastery learning in the literature necessitates a careful analysis of the results.

Several factors could have reduced the effectiveness of the mastery component. The circumstances that seem most likely and worth examining closely are a lack of generalizability or transfer of training between the format of the mastery component and the format of the evaluation component, a lack of continuity in the availability of the mastery component, student motivation issues related to engaging in the practice of taking mastery quizzes repeatedly, and the perceived or real impact that taking quizzes had on the overall grade. These issues will be examined individually and in detail.

The first issue regarding a lack of generalizability or transfer of training between the format of the mastery component and the format of the evaluation component potentially had the greatest impact on the study’s outcome. The mastery component took
the form of a multiple-choice, 10-question quiz. The questions were selected randomly from a bank of, on average, 100 questions from the chapter and lecture. The evaluation took the form of an essay exam, also approximately 10 questions, taken from the chapter and lecture. The subject matter was of course the same for the mastery quiz and the exam, but the questions themselves were not worded the same and the type of question (e.g., definition vs. conceptual application) asked on the mastery quiz was considerably different from the essay exam. As such, the type of response needed to correctly answer each kind of question was also very different and perhaps required different study techniques or different testing skills.

One conclusion is that transfer of training between these two varied presentation formats may not have occurred as effectively as assumed. In other words, studying for the mastery quizzes may not have better prepared students for an essay assessment.

Some substantiation for this proposition can be found in the student comments elicited on the survey given at the end of the semester (see Appendix K). Some of the relevant comments are as follows: “I would use the same questions on the test, in the review quiz, so it’s mastered,” “The questions that were on the quizzes were never on the test,” “Make the questions more related to the chapter review (study objectives),” and “A lot of time we were quizzed over info we didn’t need to know (for the test).” These comments point to the fact that the students were aware of the presentation format discrepancy and were even able to link this difference to a perceived decrease in the mastery component’s effectiveness. This student feedback may point to a variance in the skills needed to answer questions of different type. This variance is perhaps best characterized by Bloom’s taxonomy of educational objectives (Bloom, 1956).
Bloom’s taxonomy identifies six levels of learning behavior related to the type of responding required to answer an evaluative question. In other words, it is a way to classify the type of information being asked for in a question. The first (lowest) level is “Knowledge” and a question at this level would assess basic recall performance or memorization. The levels proceed to require more critical or analytic thought until the final level of “Evaluation.” The mastery component questions primarily were indicative of Bloom’s lowest level, questions requiring only the recall or recognition of information. Correctly answering this type of question requires a different skill set (which may require different training) than that necessary for success in answering an essay question, which often requires the application of information (Bloom’s third level), often in a novel way. In fact, some of the essay exam questions also asked for the analysis, synthesis and evaluation of information, Bloom’s highest levels.

In hindsight, it is not difficult to see the disconnect between the presentation format and skill sets required of the mastery quizzes and those required of the essay examinations. It seems this discrepancy could have played a part in the mastery quizzes’ “lack of effect” on the exam scores. Given this inconsistency, a student may have done well on the mastery quizzes, but not on the corresponding essay exam; not because that student did not understand or “master” the material, but because the definition of mastering material for a multiple-choice quiz may be different than the definition of mastering material for an essay exam. In other words, the skill set needed for successful completion of the mastery quizzes was not the same as that needed for success on the essay examinations.
Another possible limitation is the lack of contact some students had with the independent variable due in part to a lack of continuity in the availability of the mastery component. The experimental design of this study required the mastery component to be available for only half of the total units. A randomization of chapters for which the mastery quiz would be available was also necessary to account for differences in difficulty from chapter to chapter. As such, the students had a mastery quiz available for only 4 or 5 (depending on section) of the 9 units covered and the schedule was necessarily inconsistent. For example, students may have had a mastery quiz available for chapter 1, but then not for chapters 2 and 3 and then a mastery quiz was next available for chapter 4. Even though this schedule was clearly announced and recorded in the syllabus and identified on Blackboard, this discontinuity apparently led some students to reportedly find it difficult to “remember” to take the mastery quizzes.

This possibility is evidenced by several comments made on the social validity survey. Some such comments were “I would either make the quizzes every week or not at all because I always forgot about them” and “I would suggest the review quizzes be put in the announcements (on Blackboard) so students remember to take them.” It is possible that the students who did not take the quizzes to mastery may also be the students who would most benefit from continuity and a predictable structure, due to a lack of self-management skills.

Another issue worth examining is student motivation to engage in the practice of taking quizzes repeatedly. For students who took all mastery quizzes, only half in each section took them until they reached mastery (5 of 10 students in section 1 and 3 of 6 students in section 2). This practice of retaking quizzes is a normal requirement of the
mastery process but the purpose of retakes in this scenario is somewhat different than the traditional use. Retaking quizzes in PSI was required for (and thus motivated by) continuation through the course. In the current mastery application, retakes functioned only to improve the student’s individual quiz grade. Each mastery quiz was worth a potential 10 points and the score was determined by the number of questions answered correctly. If 5 questions were answered correctly, the student would receive a score of 5 points. The quiz could then be retaken as many times as the student chose and the highest score would be recorded. The differential between 5 points and 10 points appeared to be sufficient to prompt the student to take the quiz again for a higher score (only 2 of the 268 final quiz scores or less than .01% were a 5 or below). So those who took the mastery quizzes almost always continued to take them until their score was higher than a 5. But the incentive to retake the quiz when the difference was just a few points (such as going from 7 to 10), appeared to be less powerful (30 of the 268 or 11% of the final quiz scores were 6, 7, or 8).

One other issue is that some students did not attempt the mastery quiz at all for some chapters. Of the 32 participants in this study, 5 students in section 1 and 11 students in sections 2 did not take all of the mastery quizzes. These 16 students cumulatively missed 20 mastery quiz opportunities. In addition to the possibility of the student simply not remembering to take the quiz, the perceived or real impact that taking quizzes had on the overall grade may have been insufficient to support complete participation. The quizzes were worth 10 points each for about 8% of the total grade. This was about the same proportion of the grade that the reading quizzes accounted for, but both accounted
for the smallest proportion of points. Future research could examine whether increasing the value of the quizzes makes a difference in participation.

Many possible issues contributing to the outcome of this study have been identified, and as such, there are many opportunities for future research. The first issue identified, relating to the differing skill requirements between the mastery quizzes and the examinations, could be addressed in several ways. The most obvious would be to create the quizzes and the examinations in the same format. For example, both multiple-choice or both essay. This is probably the easiest way to remediate the problem, and for some classes, may be an instructionally sound choice. For an introductory psychology course however, multiple-choice examinations may not be the best evaluation of comprehension (McDaniel et al., 2007).

The skills required of essay question responding are in many ways the same skills being taught and encouraged in a science-based class (critical thinking, scientific analysis, etc.) and evaluating these skills in a manner consistent with the course content (using essay examinations) may be preferable. So if we deem that essay examinations are preferable and we want to maintain consistent format between the mastery quizzes and the exams, the only obvious alternative is to offer the mastery quizzes in an essay format. However, there are problems with this approach. First, the creation of essay quizzes is much more difficult and time consuming than multiple choice quizzes within a course management system. The reason is that many course textbooks come with test question bank software or online access and most of these banks are in multiple-choice format. If essay questions are made available, they are considerably fewer than the amount needed to create a bank large enough to draw from randomly. Therefore, the questions would
most likely need to be created by the instructor and manually added to the course management system. This would be a more effortful and time-consuming task than using the pre-existing multiple choice test banks and as such, a less likely choice for most instructors.

Secondly, grading essay mastery quiz questions would also have to be done manually, as a CMS cannot automatically account for variations in wording for essay answers. This process would most likely require several proctors or assistants for an average introductory class of more than 30 students. Many universities and most community colleges do not have the infrastructure or financial resources to support these positions. So while this scenario is possible, it is not realistic for widespread use. An alternative option may be to create a different kind of multiple-choice question for the mastery quizzes: one that maintains the ability to automate grading through a course management system, while allowing for a more “essay like” assessment of a student’s comprehension. In other words, changing the format of the multiple choice questions as opposed to the format of the quiz structure itself. Knight (2009) has devised a way to test this possibility by creating multiple-choice questions that ask for information in a way similar to an essay question. For example, such a question may ask a student to compare and contrast two different theories. Such a task would require the use of higher-order thinking skills reflective of Bloom’s fourth level. To maintain the multiple-choice format, a series of possible answers could then be given from which the student would choose. The question might ask what the student would have to do in order to answer the question correctly. The answer options may look something like this: “A. Name and define two psychological theories? B. List as many psychological theories as you can think of? C.
Name two psychological theories and tell how they are similar and how they are different? or D. Describe in depth the psychological theory you understand best?" (Knight, 2009). The idea behind this type of multiple-choice question design is that it may help the student identify and practice the skills required for successful essay test responding, without requiring an actual essay answer. This is one approach currently being investigated by Knight, but additional research is required to determine the empirical support for this type of arrangement.

The other issues recognized as potential problems could also be addressed in future research. The inconsistent contact with the independent variable due to the varied availability of the mastery quizzes could potentially be lessened by sending email reminders to the students the day a mastery quiz is made available or posting an announcement on Blackboard, as one student suggested. A different experimental model could also be considered, although there are issues of reliability and independent variable integrity to contend with for every approach. The issue of student motivation to engage in mastery quizzing could be similarly addressed through changes in methodology that would allow for comparing the effects of different point values for the quizzes.

The student comments on the end of semester social validity survey have been discussed as a way to identify areas of concern and generate possible ideas for improvement. The social validity survey data also provided a great deal of information about student attitudes and mastery quiz use trends that are important points of focus for future improvements. The feedback revealed generally positive student attitudes toward the mastery quizzes. All statements on the survey were worded in a positive orientation, such that agreeing with the statement indicated a favorable evaluation of some aspect of
the mastery quizzes. Overall, 57% of student responses were in agreement with the survey statements, 21% of responses were in disagreement with the survey statements and 23% of responses were marked undecided.

The first statement of the social validity survey asked students if they thought the mastery quizzes were beneficial. Most students felt they were; 61% either agreed or strongly agreed, 20% disagreed or strongly disagreed, and 20% were undecided (from this point forward the categories of “strongly agree” and “agree” will be combined and referred to as “agreement,” and the responses of “strongly disagree” and “disagree” will be similarly combined and referred to as “disagreement”). This indicates that though not all students took full advantage of the mastery quizzes (either by not taking all of them or by not taking all of them to mastery), the majority of students rated them as helpful. This is important because the students in this study were only given the opportunity to take mastery quizzes for half or fewer of the total units. If students thought this limited exposure was beneficial, it might be reasonable to assume that students would find the mastery quizzes even more beneficial if they were made available for all units.

When students were asked if they felt they learned the chapter material more efficiently (faster) as a result of the review quizzes, more students agreed than disagreed (49% agreed, 29% disagreed, and 22% were undecided). This is an interesting finding because it does take longer, of course, to take the quizzes and potentially quite a bit longer to take the quizzes to mastery, than if this were not a requirement of the course. That being the case, the fact that half of the students indicated that the mastery quizzes helped them save time, or learn the material more quickly, is interesting.
When students were asked if they thought they were able to learn the chapter material more effectively (better) as a result of the mastery quizzes, 51% agreed with this assertion, 24% disagreed and 24% were undecided. Much like the previous statement, this outcome is encouraging as a majority of students did in fact feel the mastery quizzes improved their comprehension of the material. However, it is interesting that the actual learning outcomes from this study do not support this perception.

When asked if the time allowed to take the mastery quizzes (20 minutes) was sufficient, an overwhelming majority of students, 78%, agreed, .05% disagreed, and 17% were undecided. Clearly, 20 minutes was enough time to answer 10 multiple-choice questions, but if the mastery quiz is changed, either in length or format or both, the time allowed for completion would have to be reevaluated.

When asked if they would have liked to have mastery quizzes available for all chapters, 54% of students said yes, 27% disagreed, and 20% were undecided. This correlates with and corroborates the student comments on this issue of consistent or increased availability of the mastery quizzes. As has been discussed, there are many reasons why having mastery quizzes for every chapter would at least make it easier for students to remember to take them and at best may actually be educationally beneficial.

The final survey statement asked if the students would like to take another course that used review quizzes. This outcome was also positive as 49% agreed, 20% disagreed, and 32% were undecided. This statement may allude to possible generalizability of the intervention, in that mastery quizzes could potentially be used for many different kinds of subjects. It appears that students perceive there would be value in a more pervasive use of
the mastery component, but such a suggestion should be viewed cautiously as this perception is not confirmed by the current results.

These social validity data should be interpreted carefully, as their generally positive direction could be a result of a novelty effect or confirmation bias or any of the other problems that plague survey data. But the fact that students rated the addition of mastery quizzes to be favorable and valuable almost 2 to 1 can be viewed optimistically for future research of this type. At the least it may reflect an acceptance of the technology being used to provide college instruction and this in itself is encouraging, as the use of technology in education will only increase. It is the hope that these social validity data may bode well for future uses of both a mastery-based methodology and an Internet-based, interactive delivery system. Though the current study’s outcomes do not support the positive educational effects from these approaches, much past research does (Bradford et al., 2007; Fredrick & Hummel, 2004; Kulik et al., 1990), and this should not be forgotten.

The subject of instructor usability is one that should not be taken for granted in educational research and one this study addressed directly. The current investigator kept track of the time spent to create, implement, and maintain it. Using a course management system such as Blackboard to assist and enhance the delivery of a “traditional” course required only a minimal amount of additional effort.

One last point to consider is that much of the time spent creating the mastery quizzes went to creating custom structure and content. However, this is not a necessary part of the process. There is a gradient of time required that depends on the way the instructor wishes to arrange their class. A traditional “out of the box” course would
require much less time and maintenance than a more “customized” course format. Many
textbooks are now offering course cartridges either complementarily or for an additional
fee, with Blackboard-ready content, including tests and quizzes. Using a more
standardized format would decrease the time required to establish on online component
considerably. Therefore, assuming most instructors are going to be required to use a
course management system in some capacity in the near future, the additional time
required to add a mastery component is probably not a barrier to its use.

While the results of this study did not indicate a positive effect on learning by the
addition of a mastery component, the goal of trying to implement mastery in a more
efficacious manner using web-based technology is still worthwhile and timely. This has
proven to be a complex endeavor, one that will require additional research before clear
answers are found. As has been thoroughly examined, there is a great deal of supportive
research in both the areas of mastery learning and web-based implementation in higher
education. The results of this study should be viewed in the context of this extensive
literature base. Following this line of research may lead to a better use of our educational
resources and, most importantly, to better learning outcomes for our students.
REFERENCES


Pear, J. (2002). Teaching and researching higher-order thinking in a virtual environment. In J. A. Chambers (Ed.), *Selected papers from the 13th International Conference on College Teaching and Learning* (pp. 143-150). Jacksonville, FL: Florida Community College at Jacksonville.


Appendix A

Human Subjects Institutional Review Board
Letter of Approval
Date: September 10, 2008

To: Eric Fox, Principal Investigator
    Amy Scrima, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number: 08-08-13

This letter will serve as confirmation that your research project entitled “Implementing a Mastery Model through Self Quizzing in an Online Learning Environment” has been approved under the expedited category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

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

Approval Termination: September 10, 2009
Appendix B

Human Subjects Institutional Review Board
Informed Consent Form
You have been invited to participate in a research project titled "Implementing a mastery model through self quizzing in an online learning environment." This project will serve as Amy Scrima's doctoral dissertation for the requirements of the Doctorate of Philosophy. This consent document will explain the purpose of this research project and will go over all of the time commitments, the procedures used in the study, and the risks and benefits of participating in this research project. Please read this consent form carefully and completely and please ask any questions if you need more clarification.

What are we trying to find out in this study?
This study will be looking at the possible benefits of mastery learning using a course management system at the undergraduate college level. We will do this by comparing student test scores with and without the use of mastery quizzes.

Who can participate in this study?
All students in sections of PSY 201 taught by Amy Scrima may have their data be part of this study. This data will be coded such that no personally identifiable information will be associated with it when it is being analyzed or reported. This means that your name will no longer be associated with your grades.

Where will this study take place?
This study will take place as part of the normal activities of class for the fall semester of 2008.

What is the time commitment for participating in this study?
There is no additional time commitment for participation in this study outside of normal course expectations.

What will you be asked to do if you choose to participate in this study?
You will not be asked to engage in any extra work to participate in this study.

What information is being measured during the study?
This study will look at unit exam, final exam and quiz scores for the fall semester of 2008.

What are the risks of participating in this study and how will these risks be minimized?
There is no discernable risk to you by allowing your data to be used in this study. A student volunteer will collect this consent form and place it in an envelope with the others. The envelope will not be opened by the student investigator until after final grades have been submitted at the end of the semester. Therefore, the instructor will not know if you chose to have your data be part of the study until after the class is finished, so your choice can not affect your grade.
What are the benefits of participating in this study?
The benefit of participation in this study is the contribution you are making toward improving the college's use of technology and instruction in the classroom. You are also helping the student investigator obtain the data necessary for the completion of her degree requirements.

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

Is there any compensation for participating in this study?
There is no monetary compensation.

Who will have access to the information collected during the study?
The analysis of the data for this study will be done by the student investigator, Amy Scrima. Grading will also be done by a teaching assistant. All personally identifiable information will be kept by Amy Scrima in a locked file cabinet in her office for 2 years. After that time all documents containing names will be destroyed. No names will be used when data are tabulated or reported.

What if you want to stop participating in this study?
You can choose to stop participating in the study at any time for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences either academically or personally if you choose to withdraw from this study.

Should you have any questions prior to or during the study, you can contact the primary investigator, Eric Fox at eric.fox@wmich.edu. You may also contact the Chair, Human Subjects Institutional Review Board at 269-387-8293 or the Vice President for Research at 269-387-8298 if questions arise during the course of the study.
Appendix C

Example of Unit Exam
Psy 201 Test
Unit 1

Name______________________________  Score_____/40

1. What is the definition of psychology (as discussed in class)? (3)

2. Compare and contrast 2 of the different perspectives in psychology. (4)

3. What is the difference between a hypothesis and a research question? (4)

4. What are the three components of the scientific method and what do they lead to? Why is research done this way? (5)

5. What is the difference between science and pseudoscience? Why is it important we be able to tell the difference? (5)

6. Why is it important to understand that all opinions are not created equal? (2)

7. What is the difference between a psychologist and a psychiatrist? (4)

8. Give examples of why you might use 3 of the 8 critical thinking guidelines. (6)

9. Discuss 2 jobs psychologists might hold and how they could impact the community. (4)

10. How might asking questions help you learn something new? (3)
Appendix D

Example of Reading Quiz
Name one kind of logical reasoning.
Appendix E

Example of Assignment
**Behavior change project** – choose a behavior you would like to change in yourself or someone else or a pet. Decide whether you want to increase or decrease the behavior and then make a three-term contingency describing your plan. Take baseline for 3 days (Wed – Fri), then implement your contingency for 5 days (Sat – Wed). Write up the results in this format:

1. Behavior
2. 3 term contingency diagram
3. Graph showing both baseline and intervention
4. Paragraph or 2 explaining how things went, any problems and plans for the future

Example –

I want to decrease my son’s pinching.

Attention → Parker pinches → no attention

Parker pinched on average 3 times a day. When he pinches, the person will immediately leave the room for 1 minute (penalty).

Graph looks like this –

```
<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinches</td>
<td>5 4 3 2 1</td>
<td>5 4 3 2 1</td>
</tr>
<tr>
<td>Days</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
</tbody>
</table>
```

Result – Parker’s pinching declines to 1 pinch per week the next week, then 0 the week after that.

I think this intervention was successful, but I will continue it as necessary until the behavior has been completely extinguished for one month.
Appendix F

Demographics Questionnaire
PSY 201 Introduction to Psychology Fall 2008
Anonymous Demographics Questionnaire

1. What is your current cumulative grade point average?
   Answer: __________

2. What is your ethnicity?
   Answer: __________

3. How many credit hours are you taking this semester?
   Answer: __________

4. What is your undergraduate student status? (circle one)
   Freshman    Sophomore    Junior    Senior

5. Are you currently employed? (circle one)
   Yes          No

6. If so, how many hours per week do you work?
   Answer: __________

7. Would you describe yourself as a traditional or non-traditional student?
   Answer: __________
Appendix G

Social Validity Survey
PSY 201
End of Semester Survey

Note: Please do NOT write your name on this survey!

Instructions: Indicate your agreement with each of the statements below by circling your answer. Use the scale below to make your choice. Then please answer the final question by writing your answer below the question.

| SD | strongly disagree |
| D  | disagree          |
| U  | undecided         |
| A  | agree             |
| SA | strongly agree    |

1. Overall, the review quizzes were beneficial
   
   SD  D  U  A  SA

2. I learned the material more efficiently when I took review quizzes (that is, I learned it faster)
   
   SD  D  U  A  SA

3. I feel I learned the material more effectively when I took review quizzes (that is, I understood it better)
   
   SD  D  U  A  SA

4. The time allowed to take the quiz was sufficient
   
   SD  D  U  A  SA

5. I would like to have had the opportunity to have review quizzes for all chapters
   
   SD  D  U  A  SA

6. I would like to take another course that used review quizzes
   
   SD  D  U  A  SA

7. What, if anything, would you change about the review quizzes?
<table>
<thead>
<tr>
<th>DATE</th>
<th>TOPIC</th>
<th>ASSIGNMENT/QUIZ</th>
<th>EXAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 3</td>
<td>Introduction</td>
<td>A: Paper</td>
<td></td>
</tr>
<tr>
<td>September 5</td>
<td>How we Learn – 1</td>
<td>A: Study skills</td>
<td></td>
</tr>
<tr>
<td>September 8</td>
<td>How we Learn – 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>September 10</td>
<td>Ch. 1</td>
<td>Reading quiz/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Review Quiz</td>
<td>Paper due</td>
<td></td>
</tr>
<tr>
<td>September 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September 17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September 19</td>
<td>Ch. 4</td>
<td></td>
<td>Exam 1</td>
</tr>
<tr>
<td>September 22</td>
<td>Ch. 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>September 24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September 26</td>
<td>Ch. 6</td>
<td></td>
<td>Exam 2</td>
</tr>
<tr>
<td>September 29</td>
<td>*Review Quiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 6</td>
<td></td>
<td></td>
<td>Exam 3</td>
</tr>
<tr>
<td>October 8</td>
<td>Ch. 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 10</td>
<td></td>
<td>A: Behavior Change</td>
<td></td>
</tr>
<tr>
<td></td>
<td>project/AD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 17</td>
<td>Ch. 10</td>
<td></td>
<td>Exam 4</td>
</tr>
<tr>
<td>October 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 24</td>
<td></td>
<td></td>
<td>Exam 5</td>
</tr>
<tr>
<td>October 26</td>
<td>Ch. 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Review Quiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October 31</td>
<td></td>
<td>Assignment due/GR</td>
<td></td>
</tr>
<tr>
<td>November 3</td>
<td></td>
<td></td>
<td>Exam 6</td>
</tr>
<tr>
<td>November 5</td>
<td>Ch. 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Review Quiz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>November 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>November 10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>November 12</td>
<td></td>
<td>Exam 7</td>
</tr>
<tr>
<td>November 14</td>
<td>Ch. 13</td>
<td>Reading quiz</td>
</tr>
<tr>
<td>November 17</td>
<td></td>
<td>AD</td>
</tr>
<tr>
<td>November 19</td>
<td></td>
<td>GR</td>
</tr>
<tr>
<td>November 21</td>
<td></td>
<td>Exam 8</td>
</tr>
<tr>
<td>November 24</td>
<td>Activity</td>
<td></td>
</tr>
<tr>
<td>November 26</td>
<td></td>
<td>No Class</td>
</tr>
<tr>
<td>November 28</td>
<td></td>
<td>No Class</td>
</tr>
<tr>
<td>December 1</td>
<td>Ch. 14</td>
<td>Reading quiz</td>
</tr>
<tr>
<td></td>
<td>*Review Quiz</td>
<td></td>
</tr>
<tr>
<td>December 3</td>
<td></td>
<td>A: Review project/AD</td>
</tr>
<tr>
<td>December 5</td>
<td></td>
<td>GR</td>
</tr>
<tr>
<td>December 8</td>
<td></td>
<td>Exam 9</td>
</tr>
<tr>
<td>December 10</td>
<td>Final exam prep</td>
<td>Assignment due</td>
</tr>
<tr>
<td>December 12</td>
<td>Final exam</td>
<td>Final exam</td>
</tr>
</tbody>
</table>

A= Assignment  
AD= Activity or demonstration  
GR= Group work and review
Appendix I

Sample IOA Calculation Sheet
IOA Calculation Sheet

Chapter 1

1. __ __
2. __ __
3. __ __ __ __ __
4. __ __ __ __ __ __
5. __ __
6. __ __

A =

D =

IOA =

7. __ __ __ __ __
8. __ __ __ __
9. __ __
10. __ __ __ __
11. __ __ __
12. __ __ __
Appendix J

Graphs of Individual Student Averaged Exam Scores by Condition
Participant 22

Score

Exam

Participant 23

Score

Exam

Participant 24

Score

Exam

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix K

End of Semester Survey Student Responses to Question 7, "What if anything would you change about the review quizzes?"
Section 1

"Maybe make them the objective questions or more like the objective questions. They didn't really help me study."

"I feel the review quiz was a great idea and very helpful to me in class. And it also helped me to understand the chapter more better that I had read."

"Leave them up for reviewing after the time expires so that we can go back and look a them for the final or for reference/practice. Once the grade is recorded after time allowed is up, grade can't be changed but review quizzes can still be taken."

"Less of them."

"I wouldn't do the quizzes because they didn't help."

"I would use the same questions on the test in the review quiz so its mastered."

"Pertain a little more closely to the test, overall really good and beneficial."

"Nothing"

"Nothing, really"

"More make up time for it"

"I would either make the quizzes every week or not at all because I always forgot about them."

"Some of the questions were more from the book than what we discussed in class. I would suggest also that the review quizzes be put in the announcements so students remember to take them. We need more review quizzes, they really do help."

Section 2

"Nothing"

"I believe that they should be offered but not mandatory"

"I would like them for all the chapters and maybe center them more around the study objectives sheet"

"Every chapter, get rid of reading quiz"

"The questions that were on the quizzes were never on the test"
“Nothing”

“None”

“I don’t think they benefited because the questions we did not go over much and were never on the test.”

“I think if the questions were shorter, they would have been easier to understand.”

“I think it would be better if the questions didn’t include examples like, “Sally took the test…” Being timed and really having to think about those stressed me out.”

“I would add questions that focused on the lecture portion of the class, since some of the exam questions were based on lecture definitions.”

“The review quizzes definitely caused me to read the chapter and know where to find the answers.”

“Make the questions more related to the chapter review. A lot of time we were quizzed over info we didn’t need to know”
Appendix L

Tables and Figures
**Table 1**

*Condition Assignment Table*

<table>
<thead>
<tr>
<th></th>
<th>Unit 1</th>
<th>Unit 2</th>
<th>Unit 3</th>
<th>Unit 4</th>
<th>Unit 5</th>
<th>Unit 6</th>
<th>Unit 7</th>
<th>Unit 8</th>
<th>Unit 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sec. 1</td>
<td>M</td>
<td>N</td>
<td>M</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>M</td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>Sec. 2</td>
<td>N</td>
<td>M</td>
<td>N</td>
<td>N</td>
<td>M</td>
<td>M</td>
<td>N</td>
<td>M</td>
<td>N</td>
</tr>
</tbody>
</table>

*Note.* M = mastery condition; N = non-mastery condition
Table 2

*Mean Exam Scores by Condition for Section 1*

<table>
<thead>
<tr>
<th></th>
<th>Unit Exam</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 (n = 15)</td>
<td>4 (n = 15)</td>
<td>5 (n = 15)</td>
<td>8 (n = 15)</td>
<td></td>
</tr>
<tr>
<td>Non Mastery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M$</td>
<td>32.13</td>
<td>30.13</td>
<td>30.2</td>
<td>32.33</td>
<td></td>
</tr>
<tr>
<td>$SD$</td>
<td>6.36</td>
<td>7.08</td>
<td>5.99</td>
<td>5.72</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>23-40</td>
<td>15-40</td>
<td>18-40</td>
<td>24-40</td>
<td></td>
</tr>
<tr>
<td>Mastery</td>
<td>1 (n = 15)</td>
<td>3 (n = 15)</td>
<td>6 (n = 15)</td>
<td>7 (n = 15)</td>
<td>9 (n = 15)</td>
</tr>
<tr>
<td>$M$</td>
<td>25.93</td>
<td>32.33</td>
<td>33.73</td>
<td>28.93</td>
<td>29.33</td>
</tr>
<tr>
<td>$SD$</td>
<td>8.77</td>
<td>5.86</td>
<td>5.66</td>
<td>8.34</td>
<td>7.0</td>
</tr>
</tbody>
</table>

*Note.* The maximum possible score for each exam was 40.
Table 3

*Mean Exam Scores by Condition for Section 2*

<table>
<thead>
<tr>
<th>Condition</th>
<th>1 (n = 17)</th>
<th>3 (n = 17)</th>
<th>6 (n = 17)</th>
<th>7 (n = 17)</th>
<th>9 (n = 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non Mastery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M)</td>
<td>26.53</td>
<td>29</td>
<td>33.82</td>
<td>28.65</td>
<td>26.76</td>
</tr>
<tr>
<td>(SD)</td>
<td>9.32</td>
<td>5.16</td>
<td>6.55</td>
<td>7.98</td>
<td>7.62</td>
</tr>
<tr>
<td>Mastery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M)</td>
<td>29.59</td>
<td>28.29</td>
<td>29.12</td>
<td>34.47</td>
<td></td>
</tr>
<tr>
<td>(SD)</td>
<td>9.82</td>
<td>6.53</td>
<td>7.18</td>
<td>4.12</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>14.5-34</td>
<td>13-35</td>
<td>10-33</td>
<td>17-35</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The maximum possible score for each exam was 40.
Table 4

*Mastery Quiz Attempts by Unit for Sections 1 and 2*

<table>
<thead>
<tr>
<th>Section 1</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 15)</td>
<td>(n = 15)</td>
<td>(n = 15)</td>
<td>(n = 14)</td>
<td>(n = 14)</td>
</tr>
<tr>
<td>M</td>
<td>4.9</td>
<td>3.5</td>
<td>4.8</td>
<td>3.3</td>
<td>3.4</td>
</tr>
<tr>
<td>SD</td>
<td>3.2</td>
<td>3.2</td>
<td>4.5</td>
<td>3.4</td>
<td>4.9</td>
</tr>
<tr>
<td>Range</td>
<td>1-12</td>
<td>1-11</td>
<td>2-19</td>
<td>1-13</td>
<td>1-20</td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section 2</th>
<th>2</th>
<th>5</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 15)</td>
<td>(n = 9)</td>
<td>(n = 12)</td>
<td>(n = 14)</td>
</tr>
<tr>
<td>M</td>
<td>6.7</td>
<td>2.1</td>
<td>2.5</td>
<td>3.9</td>
</tr>
<tr>
<td>SD</td>
<td>6.0</td>
<td>2.9</td>
<td>2.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Range</td>
<td>2-20</td>
<td>1-9</td>
<td>1-11</td>
<td>1-10</td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 1. Unit exam score averages by section and condition.
Figure 2. Graph of the one student’s exam scores that reflects a clear difference between mastery and non-mastery conditions. This is for participant 20 in section 2.
Figure 3. Graphs of two students’ exam scores that reflect a difference between mastery and non-mastery conditions for all units but one. These are participants 2 in section 1 and 21 in section 2, respectively.
Figure 4. Graph of most representative student exam score distribution between mastery and non-mastery conditions. This is participant 10 in section 1.
Figure 5. Exam grade distribution by condition for section 1.
Figure 6. Exam grade distribution by condition for section 2.
Figure 7. Exam means for students who took all quizzes to mastery in sections 1 and 2 combined.
Figure 8. Student responses to question 1, “Overall, the review quizzes were beneficial.”
Figure 9. Student responses to question 2, "I learned the material more efficiently (faster) when I took review quizzes."
Figure 10. Student responses to question 3, “I learned the material more effectively (better) when I took review quizzes.”
Figure 11. Student responses to question 4, “The time allowed to take the quiz was sufficient.”
Figure 12. Student responses to question 5, “I would like to have had the opportunity to have review quizzes for all chapters.”
Figure 13. Student responses to question 6, “I would like to take another course that uses review quizzes.”