# THE CLICKER TECHNIQUE: PROMOTING LEARNING AND GENERALIZATION WHILE CONSERVING TEACHING TIME

by

LINDSAY SENIOR ANDERSON

B.S., Colorado State University, 2007

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

Of the requirement for the degree of

Master of Arts

Department of Psychology and Neuroscience

This thesis entitled: The Clicker Technique: Promoting Learning and Generalization while Conserving Teaching Time Written by Lindsay Senior Anderson Has been approved for the Department of Psychology and Neuroscience

Dr. Alice F. Healy

Dr. Lyle E. Bourne, Jr.

Dr. Matthew C. Jones

The final copy of this thesis has been examined by the signatories, and we Find that both the content and the form meet acceptable presentation standards Of scholarly work in the above mentioned discipline.

IRB protocol # <u>0708.15</u>

Anderson, Lindsay Senior (Department of Psychology and Neuroscience)

The Clicker Technique: Promoting Learning and Generalization while

Conserving Teaching Time.

Thesis directed by College Professor of Distinction Alice F. Healy

#### Abstract

The *clicker* technique is a newly developed system that uses frequent testing in the classroom to enhance students' understanding and provide feedback to them. Under the clicker technique, instructors can use the performance of a class on clicker questions to determine whether or not information covered by the clicker questions needs further teaching, thus presenting itself as a potential method of conserving teaching time by dropping information known by a large portion of a group from future teaching time. Three experiments compared fact learning under the clicker technique, via its tendency to compress teaching time and its partially individualized instruction, to fact learning under other repeated testing possibilities, such as dropout and full-study procedures. Experiment 1 explored initial fact acquisition under the clicker technique, Experiment 2 explored the durability of knowledge acquired under the clicker technique on both immediate and delayed tests, and Experiment 3 explored the durability and generalizability of knowledge acquired under the clicker technique on both immediate and delayed tests. Overall, results support the clicker technique as a viable method for promoting efficient and generalizable learning while compressing teaching time without sacrifice of amount learned.

CHAPT	ER	
I.	INTRODUCTION	.6
II.	EXPERIMENT 11	2
	Method1	3
	Results	16
	Discussion1	18
III.	EXPERIMENT 2	20
	Method2	23
	Results	26
	Discussion	29
IV.	EXPERIMENT 3	32
	Method	33
	Results	39
	Discussion	16
V.	GENERAL DISCUSSION	50
REFERI	ENCES5	;8
APPENI	DIX	
	A. EXPERIMENTS 1 AND 2 STIMULI	51
	B. EXPERIMENT 3 STIMULI	52

## FIGURES

Figure

17	Mean proportions of correct recall for Block 1 and Block 4 as a function of learning condition in Experiment 1	1.
27	Mean proportions of correct recall for Block 1 and Block 4 as a function of learning condition in Experiment 2	2.
	Interaction between test time and condition in Experiment 2	3.
29	Forgetting from the immediate test to the retention test as a function of condition in Experiment 2	4.
40	Interaction at test between learning format and condition in Experiment 3	5.
41	Interaction at test between test time and condition in Experiment 3	6.
42	Three-way interaction at test between test time, learning fact format, and condition in Experiment 3	7.
43	Three way interaction at test between test time, learning fact format, and test fact format in Experiment 3	8.
44	Forgetting between the immediate test and the retention test as a function of condition in Experiment 3	9.
45	0. Interaction between test time and item difficutly in Experiment 3	10
46	1. Three-way interaction between item difficutly, learning fact format, and test fact format in Experiment 3	11

#### CHAPTER I

#### INTRODUCTION

The future of our society depends largely on the quality of education that children and young men and women receive in the classroom. In modern university and educational settings, learning is often taking place in the context of large classes. Large classes indicate that more people are getting an education, but as a consequence instructors may easily lose a sense of how well students are understanding and following the material being taught. In addition to large class sizes, students are held responsible for learning large amounts of information. The current research helps to identify conditions that permit compression of training, that facilitate long-term retention of learning, and that facilitate the transfer of knowledge, thus accelerating learning and enhancing training efficiency. Such conditions can be applied most immediately to classroom settings and ultimately to relevant job training settings.

The purpose of the current research is to explore a newly developed classroom teaching procedure, the clicker technique, which uses frequent testing as a way to provide immediate feedback to instructors about students' understanding of material (Dreifus, 2005). There are other systems similar to the clicker technique, which are often referred to as personal response systems (PRS), wireless response systems (WRS), electronic voting systems (EVS), electronic response systems (ERS), electronic polling systems (EPS), and classroom communication systems (CCS), but all of these systems operate according to the same principles. The use of the clicker technology, and essentially all of these systems, involves instructors giving periodic multiple-choice probe questions to students, who then respond via a hand held device. The device used in the clicker technique is an "*i*clicker," which has buttons labeled A through E. Instructors receive immediate feedback describing which questions were missed and the

distribution, which is available to the class, of how many students selected specific answer choices. The results of these tests aid the instructor in deciding whether or not to spend additional time on certain material and in identifying and understanding where confusion might be occurring. The clicker technique has been implemented internationally on approximately 700 higher-level education campuses and in some K-12 classrooms. Currently, there is no standard method of how to use the clicker system most effectively, and most of the research on these student response systems takes place in classroom settings.

The current literature on response systems in classrooms reveals that students and instructors have an overall positive opinion of the method and its effectiveness. The clicker technique encourages and facilitates student participation by providing a secure environment for students to answer in-class questions because students' responses are anonymous to the class (Stuart, Brown, & Draper, 2004). The anonymity of the system eliminates the risk of embarrassment when answering questions, thus students tend to participate more willingly and frequently.

The effectiveness of the clicker system in the classroom also seems to depend on students' class standing and attitudes as well as instructors' experience teaching with the clicker technology. Trees and Jackson (2007) found that the effectiveness of the technology relied on students accepting the system's potential to have a positive impact on their learning, with lower-division university-level students being more accepting than upper-division university level students. Students who are educated about the value of feedback for learning and who are already interested in being involved in the class tend to benefit more from the technology than students who do not know about the value of feedback and who are not interested in class involvement (Trees & Jackson, 2007). Duncan (2005) advocates that instructors inform students

as to why they are using the clicker technique, before use has begun, in order for students to accept the system as a potentially valuable learning tool. Although students have reported that the clicker system helped to increase their understanding of material (Greer & Heaney, 2004), the clicker system's positive effect on learning is not a given. As instructors gain experience and flexibility in incorporating the clicker technique into their lectures, students' learning benefits increase (Draper & Brown, 2004; Duncan, 2005). The latter finding highlights the importance of identifying conditions under which the clicker technique is most effective, so that instructors can be trained on how to use the clicker technique to maximize learning.

The literature on the clicker technique (and related systems) in the classroom has also pointed towards a positive correlation between clicker use and learning outcomes. Kennedy and Cutts (2005) found a positive correlation between learning outcomes and the proportion of clicker questions answered correctly during class. A causal relationship between clicker use and learning outcomes cannot be drawn from the latter experiment, but the finding may indicate that correctly answering questions in class could have an influence on later learning outcomes. One explanation for this relationship was offered by Mayer et al. (2008), who found that a class that received clicker questions got 1/3 of a grade point higher in the course than sections of the same class that did not receive clicker questions. The class that received clicker questions was more cognitively engaged, thus, according to a generative theory of learning (Wittrock, 1989), it should perform better than the more passive learning classes without clicker questions (Mayer et al., 2008). Donovan (2008) drew a more direct link between clicker questions and learning outcomes by showing improvement between an in-class concept question and performance on a corresponding exam question. Such classroom data are valuable, but neither a causal connection between the clicker technique and later learning outcomes nor an explanation of what

components of the technique enhance or diminish this connection can be drawn from these studies.

Two laboratory experiments were conducted by Campbell and Mayer (2009), which provided preliminary evidence in favor of a questioning effect, which states that students learn better when they answer questions and get feedback during college lectures than when they are presented with the same information in a traditional lecture style. Overall, the clicker technique provides students with multiple opportunities to actively recall and use information that they are being taught, which seems to be one of the most fundamental aspects of the technique's success.

Sometimes clicker questions are coupled with in-class discussion. Students' attention can diminish during traditional lectures, but clickers can be used as a tool to keep students engaged and active in the learning process by stimulating peer discussion through asking conceptual questions (Duncan, 2005). In a study by Smith et al. (2009) with the clicker system, it was found that using the clicker system during class discussion enhanced students' understanding of material, even when students had no prior knowledge of the covered material before the discussion period. The latter experiment focused on the role of discussion in the enhancement of learning, while using the clicker system to provide questions to assess learning. Rather than using the clicker system to assess the effects of discussion, the current research examines the clicker technique as a possible way to shorten (i.e., conserve or compress) classroom teaching time without any sacrifice of the amount learned.

The clicker technique increases the amount of time used for testing in the classroom. This additional time expenditure might appear to be a disadvantage if, as commonly assumed, learning takes place primarily while new material is presented and studied and testing is primarily a tool for assessing learning. Contrary to this assumption, research has demonstrated

that testing has benefits above and beyond its use in assessment (e.g., Carpenter & DeLosh, 2005). Indeed, repeated testing produces greater retention of learned material than does repeated studying (Karpicke & Roediger, 2008), and practicing retrieval of information increases the rate of learning while reducing the rate of forgetting as compared to repeated studying (Carpenter, Pashler, Wixted, & Vul, 2008). The testing effect underlies the clicker technique, such that students' learning might be enhanced through in-class testing because the students are provided with additional testing opportunities.

The *dropout procedure*, introduced by Rock (1957), is a laboratory model resembling the clicker technique. In the dropout procedure, items mastered on a learning trial are not represented or retested (Karpicke & Roediger, 2008; Rock, 1957). Performance under the dropout procedure is often no worse than under full study, indicating that time spent on known items is not necessary for improved learning. The dropout procedure, therefore, is one method that can be used to compress teaching time (Pyc & Rawson, 2007). Under the clicker technique, instructors usually drop from further teaching material that the class has mastered, as indicated by the class's responses to clicker questions. Like the dropout procedure, the clicker technique provides a method of conserving study time. Given this similarity between the dropout and clicker procedures, it is possible to use a laboratory task to examine some of the putative classroom advantages of the clicker technique.

The clicker technique resembles the dropout procedure, such that instructors drop from further discussion material that the class has mastered, as indicated by the class's responses to probe questions. Given the similarities between the dropout procedure and the clicker technique, it is possible that the clicker technique may also be an effective method of teaching compression.

Experiment 1 established the validity of the clicker technique as being an effective method of fact acquisition while compressing teaching time. Experiment 2 focused on replicating the results obtained in Experiment 1 and examined the durability of knowledge learned under the clicker technique. Experiment 3 examined the generalizability of knowledge acquired under the clicker technique and also further examined the efficiency of learning and durability of knowledge learned under the clicker technique the clicker technique.

#### CHAPTER II

#### **EXPERIMENT 1**

The purpose of the first experiment was to determine how a laboratory analogue of the clicker technique compared to other repeated testing possibilities in a fact-learning experiment. Our laboratory model of the clicker technique resembles classroom clicker use, such that only questions missed by a large portion of a group remained in future teaching rounds. The inclusion in future teaching rounds of questions missed by a large portion of a group mimics the additional time instructors spend on material following a clicker question. The issue of interest in Experiment 1 was whether it was beneficial to continue to study and be tested over already mastered material, and if that material was specific to individuals.

In Experiment 1, participants learned a set of generally unknown facts across four teaching rounds. We compared six conditions, all of which required full study in Rounds 1 and 4, but five of which varied the facts presented during Rounds 2 and 3. We compared a full-study condition, which contained all facts in all teaching rounds, with five other conditions, dropout, yoked, and three clicker conditions, which varied in terms of the requirements that determined the facts presented during Rounds 2 and 3. These five conditions all involved study-time compression and were roughly the same in the amount of compression. Compression in the dropout condition was based on the performance of the tested participant, in the yoked condition on the performance of another participant, and in the clicker conditions on the performance of a group of participants. On the basis of previous results using the dropout procedure, it was predicted that improvements from Round 1 to Round 4 would be no worse in the dropout and clicker conditions than in the full-study condition, despite the reduced study time, thus demonstrating that extra study does not necessarily improve performance. If there was no

benefit for the full-study condition, then there would be evidence supporting an effective technique of study-time compression. It was also predicted that improvements from Round 1 to Round 4 would be greater in the dropout condition than in the yoked and clicker conditions, demonstrating an advantage for individualized studying. Individualized studying for course instruction has been investigated in the past. For example, the Personalized System of Instruction is an effective computerized program consisting of individual lesson plans and feedback (Keller, 1974). This system, however, has no effect on student study time in courses (Kulik, Kulik, & Cohen, 1979), and it cannot be implemented in large classes due to its highly individualized nature. The clicker technique might be an efficient method for gauging the understanding of a large number of people at once so that teaching time may be adjusted accordingly.

#### Method

#### **Participants**

Seventy-two undergraduate University of Colorado students participated in order to fulfill partial requirements for an introductory level psychology course. Participants included 25 men and 47 women.

#### Design

The design was a 6 (condition) X 2 (Round 1 vs. Round 4) mixed factorial, with the first factor manipulated between subjects. The first variable was condition. There were four study-test rounds per experimental session. The first and last rounds included all 64 facts for all conditions. The facts presented in the second and third rounds varied across conditions. The conditions are the (a) full-study, (b) dropout, (c) yoked, (d) Clicker 25, (e) Clicker 40, and (f) Clicker D conditions. For the full-study condition participants studied and were tested over all of

the facts in every round. In the dropout condition, mastered facts were dropped from subsequent rounds. Participants in the yoked condition saw only the facts that a matched participant in the dropout condition saw. The Clicker 25 condition was a laboratory analogue of how the clicker technique is used in the classroom, such that participants saw only the facts that were missed by greater than 25% of participants in the full-study condition on the previous round. Participants in the Clicker 40 condition saw only the facts that were missed by greater than 40% of participants in the full-study condition on the previous round. Finally, the Clicker D condition saw the average number of facts presented in Rounds 2 and 3 of the dropout and yoked conditions. The facts selected for presentation in Round 2 were the top 41 missed facts in Round 1 of the full-study condition, and the facts presented in Round 3 were the top 28 missed facts in Round 2 of the full-study condition. The remaining variable, learning round, was manipulated within subjects. The dependent variable was the proportion of correct responses.

The clicker conditions constitute a laboratory analogue of how response clickers are used by instructors in the classroom, such that all of the participants in the clicker conditions were tested over material that was missed by a portion of participants in the full-study condition. Testing of facts was done via cued recall rather than multiple choice in order to minimize the effects of guessing. Responses were hand scored so that responses that were incorrect due only to misspellings were counted as correct. The facts presented to participants in both the yoked and the clicker conditions were dependent on the items presented in the dropout and the full study conditions, respectively. This control allowed for comparisons of the yoked condition to the dropout condition and of the clicker conditions to the full-study condition. These comparisons helped provide evidence for individualized learning and learning time compression. It should be noted, however, that this method meant that participants in the yoked condition were necessarily

tested after those in the dropout condition, and participants in the clicker conditions were necessarily tested after those in the full-study condition. With those constraints, participants were assigned to conditions by a fixed rotation based on the time of arrival for testing. *Materials* 

The current study utilized a version of a fact-learning task adapted from a study by Kole and Healy (2007). A set of 64 true facts was selected from this fact-learning task. Participants learned eight true facts about eight different countries. One fact about each of eight dimensions was associated with each of the eight countries, with a total of one fact per dimension for every country (see Appendix A). The facts were presented as sentences, each including a country, verb phrase, and a dimension-relevant fact (e.g., "Malawi's citizens speak *Chichewa*"). In the example, the italicized word was what participants would recall at test. Each fact for all dimensions was a one-two word answer.

#### Procedure

Participants were tested individually on Apple i-Mac computers. Participants were informed that they would be viewing several sets of eight facts about different countries and that they would be tested on their ability to recall those facts. The experimenter then initiated the presentation of the fact lists. Each fact was presented individually for 3 s in blocks of eight facts, with one fact per country in each block. Each of the eight fact dimensions was used once in each block as was each of the eight countries. The order of the presentation and testing of the facts in a block were different so that participants could not use a serial order mnemonic strategy. After each block, participants were given a cued recall test over the facts presented within the block. Participants were given the country name and a verb phrase, followed by a blank (e.g. Malawi's citizens speak \_\_\_\_\_\_) and were to fill in the appropriate answer. Participants were given 9 s to

begin a response before the program would automatically proceed to the next question. There were four rounds, each consisting of eight blocks. The same eight facts occurred within each block, but the order in which they were presented was random within the blocks. In the dropout, yoked, and clicker conditions, in which not all of the facts were included in Rounds 2 and 3, facts were still presented randomly within their respective blocks. After participants completed all four study-test rounds, they were given a debriefing form explaining what they did in the experiment, and were given course credit for their participation.

#### Results

A 6 (condition) x 2 (Block 1 vs. Block 4) ANOVA was employed. Overall, the analysis of accuracy revealed a main effect of condition, F(5, 66) = 2.36, MSE = .31, p = .0499, and a main effect of learning round, F(1, 66) = 40.64, MSE = .03, p < .0001, with accuracy being higher in Round 4 (M = .50) than in Round 1 (M = .43). Importantly, condition significantly interacted with learning round, indicating that performance from Round 1 to Round 4 differed between conditions, F(5, 66) = 2.66, MSE = .03, p = .0301 (see Figure 1). Learning from Round 1 to Round 4 was evident in all conditions but the yoked condition (M = .38 averaged across Rounds 1 and 4). A Fisher's PLSD test revealed that the Clicker 25 condition (M = .51) showed no significant difference in performance during learning in Round 1 and Round 4 (averaged) from the dropout (M = .48), the full-study (M = .53), and the Clicker D (M = .50) conditions. The Clicker D condition showed no significant difference in performance during learning in Rounds 1 and 4 from the Clicker 25, Clicker 40, full-study, dropout, or yoked conditions. The yoked condition was significantly worse than the Clicker 25, Clicker D, and full-study conditions. The Clicker 40 condition (M = .40) was the only clicker condition that performed significantly worse than the full study condition in these rounds.



*Figure 1*. Mean proportions of correct recall for Block 1 and Block 4 as a function of learning condition in Experiment 1. Error bars represent +/- 1 standard error of the mean.

An ANOVA, restricted to Round 4 of learning, revealed a main effect of condition, F(5, 66) = 2.83, MSE = .21, p = .0223. A Fisher's PLSD test on an ANOVA that was restricted to Round 4 was conducted and revealed that, as predicted, final performance in Round 4 of the full-study condition (M = .59) was not substantially better than performance in Round 4 of the dropout (M = .54), the Clicker 25 (M = .54), or the Clicker D (M = .53) conditions. The yoked condition performed the worst in Round 4 (M = .38), and the Clicker 40 condition showed intermediate performance on Round 4 (M = .42). The Fisher's PLSD test revealed that the yoked condition performed significantly worse in Round 4 than the Clicker 25, the Clicker D, the dropout, and the full-study conditions. The test also revealed that one of the clicker conditions, the Clicker 40 condition, performed significantly worse than the full-study condition.

Difference scores were computed in order to assess improvement between Round 1 and Round 4. The analysis of the difference scores revealed a significant effect of condition on improvement in accuracy, F(5, 66) = 2.66, MSE = .07, p = .0301. Importantly, there was no difference in improvement between the full-study condition (M = .12), the dropout condition (M

= .11), and the Clicker D condition (M = .07). There was neither a difference in improvement between the Clicker 25 condition (M = .06) and the full-study, dropout, or the yoked conditions nor a difference between the Clicker 40 condition (M = .06) and the full-study, dropout, or the yoked conditions. The yoked condition showed no improvement from Round 1 to Round 4 (M = 0.00). A Fisher's PLSD test was conducted, with an  $\alpha$  of .05, which revealed that the improvement of the yoked condition from Round 1 to Round 4 was significantly worse than the improvement of the Clicker D, the dropout, and the full-study conditions.

#### Discussion

Experiment 1 marked the starting point for beginning to address the question of how to most effectively use the clicker technique. By directly comparing a laboratory analogue of the clicker technique to other repeated testing possibilities, Experiment 1 bridged a gap between the existing literature on testing and the clicker technique. Results support the hypothesis that the full-study condition would not perform significantly better than the dropout or the clicker conditions (except the Clicker 40 condition). Previous findings showing no difference between dropout and full-study (Karpicke & Roediger, 2008; Rock, 1957) were replicated. Importantly, the current results show no improvement advantage of full-study over the clicker conditions. Although these are null effects that need to be interpreted with caution, taken together, they indicate that extra study over mastered material is not necessary for improved performance, thus presenting the clicker technique as an effective method of learning time compression. The implications of these results for teaching are that the clicker technique helps guide instructors towards material that reliably needs further explanation or study, compressing teaching time by eliminating time spent on material known by most students.

The results also support the second hypothesis that improvements in the yoked condition would be smaller than those in the dropout and the clicker conditions. Not only did the yoked condition perform substantially worse in Round 4 than the full-study, dropout, Clicker 25, and Clicker D conditions, it also showed no improvement from Round 1 to Round 4. The lack of improvement of the yoked condition indicates that it is of no benefit for an individual to review material not specific to his or her needs. The clicker conditions showed more improvement than the yoked condition and slightly (but not significantly) less improvement than the dropout condition. It is unrealistic and inefficient to tailor lecture time to every individual in a large class; therefore, the clicker technique is a promising method that seeks to find a middle ground by assessing the understanding of a group in order for instructors to determine whether to spend additional lecture time on material. Overall, this experiment provided support for the clicker technique as an efficient and effective method for fact learning.

#### CHAPTER III

#### **EXPERIMENT 2**

Experiment 1 established that the clicker technique is effective for efficient initial fact acquisition, but one of the goals of learning is long-term retention of knowledge. The latter is especially relevant in education, where long-term learning is one of the ultimate goals. An individual's level of performance on a given skill during learning is not necessarily indicative of how well that skill will be retained over time. Research on learning has shown that conditions that increase an individual's performance during learning often decrease performance on delayed tests, whereas conditions that decrease an individual's performance during learning often increase performance on delayed tests (Healy & Bourne, 1995; Schmidt & Bjork, 1992; Schneider, Healy, & Bourne, 2002). Conditions that enhance retention by introducing some appropriate level of difficulty during learning are said to have desirable difficulties, a term introduced by Bjork (1994). Given that conditions that decrease performance during learning are sometimes those that promote knowledge durability, instructors are presented with a predicament: They must teach a fixed amount of material in a finite amount of time, but conditions that promote durable learning (i.e., those that decrease performance during learning) might make it difficult to gauge the effectiveness of their teaching. How, then, can instructors introduce sufficient difficulty during learning to promote retention while also considering the understanding of the class?

Retrieval practice, or testing, is a well-established method of increasing difficulty during learning that enhances long-term retention (Bjork, 1994; Roediger & Karpicke, 2006). Various schedules of repeated testing during learning have been proposed to be more effective than others for promoting retention. Specifically, expanding interval retrieval practice is a method of

increasing difficulty during learning that involves introducing intervals, which contain intervening study items, of expanding lengths between studying and testing of to-be-learned items (Schmidt & Bjork, 1992). Expanding retrieval practice has shown to degrade performance during acquisition but to enhance retention performance relative to constant intervals, which enhance initial learning performance but degrade retention performance (Landauer & Bjork, 1978); However, the benefits of expanding retrieval practice have been called into question with results demonstrating no differences between expanding and constant interval practice schedules (Karpicke & Roediger, 2010) and with results demonstrating that constant interval practice schedules produce superior long-term retention and that expanding interval practice schedules produce superior short-term retention (Karpicke & Roediger, 2007). Despite the controversial benefits of interval testing schedules, the absolute spacing between testing events does contribute to learning difficulty and retention performance. The well-established spacing effect (Melton, 1967) shows that increasing the absolute spacing between testing events produces better longterm retention performance (Hogan & Kintsch, 1971; Schmidt & Bjork, 1992). Contrary to the spacing effect, the clicker, dropout, and yoked conditions compress space between testing events, which, according to the spacing effect, should degrade retention performance.

In addition to compressing the absolute spacing between testing events, the clicker, dropout, and yoked conditions decrease the number of items that must be studied and tested across learning rounds. The list-length effect (Gronlund & Elam, 1994; Strong, 1912) shows that as the number of to-be-learned items in a list increases, performance declines. The list-length effect would predict that the clicker, dropout, and yoked conditions should perform better than the full-study condition because the number of to-be-learned items is reduced across learning in these conditions, relative to the full-study condition. The spacing and list-length effects make

contradictory predictions because the absolute spacing between tests is reduced (which should hurt retention) via a reduction in the number of items (which should enhance retention) across learning in the clicker, dropout, and yoked conditions.

The only difference between the clicker, dropout, and yoked conditions rests in the particular items chosen for presentation. The clicker technique is unique in that it uses the performance of a group to determine the material that should remain in further teaching, or learning, rounds. The performance of a group on a set of items may more reliably identify items that are, on average, more difficult than might the performance of an individual, as in the dropout procedure. Perhaps the testing of *particular* items, either identified as difficult by a group of people, as in the clicker technique, or by an individual, as in the dropout procedure, introduces adequate difficulty during learning to lead to knowledge durability, despite the reduced spacing between testing events. The clicker technique, therefore, might promote durability because it increases difficulty during learning via testing with the added benefit of conserving teaching time by providing instructors with a reliable indication of what material can be dropped from further teaching, spending teaching time only on material that needs further elaboration or practice.

The purpose of Experiment 2 was to explore the durability of knowledge acquired with the clicker technique. Specifically, Experiment 2 was designed to determine if the results found in Experiment 1 would hold over a 1-week retention interval. Experiment 2 also differed from Experiment 1 in that Experiment 2 included a pre-test, which occurred prior to the four study-test rounds, and an immediate post-test, which occurred after the four study-test rounds. Both the pre-test and post-test provided no opportunity for prior study. The pre-test, the post-test, and the retention test were equivalent in that they all included all 64 facts without the opportunity for prior study.

Based on the results of Experiment 1, it was again predicted that the dropout and clicker conditions would show no less improvement from Round 1 to Round 4 than the full-study condition, and it was also predicted that the dropout and clicker conditions would perform no worse than the full-study condition on the immediate post-test and the retention test. This expectation was derived from previous results on the dropout procedure (Karpicke & Roediger, 2008; Rock, 1957) and from the results of Experiment 1, which indicated that the dropout and clicker conditions performed no worse in terms of improvement from Round 1 to Round 4 than the full-study condition. Additionally, it was predicted that the yoked condition would show less improvement from Round 1 to Round 4 than the dropout, clicker, and full-study conditions and that the yoked condition would perform worse than the dropout, clicker, and full-study conditions on the immediate post-test and the retention test. This expectation was derived from the results of Experiment 1, which showed that the yoked condition showed significantly worse improvement than the dropout, Clicker D, and full-study conditions. In summary, it was predicted that the results from Experiment 1 would be replicated in Experiment 2, and that these results would also hold true across a 1-week retention interval.

#### Method

#### **Participants**

Forty-six undergraduate University of Colorado students participated in order to fulfill partial requirements for an introductory level psychology course. Participants included 22 men and 24 women. Participants were assigned to conditions by a fixed rotation based on the time of arrival for testing, as in Experiment 1, with those in the clicker conditions assigned after testing was completed in the dropout and full-study conditions and each participant in the yoked condition tested immediately after the matched participant in the dropout condition.

#### Design

The design for learning was a 4 (condition) X 2 (Round 1 vs. Round 4) mixed factorial, with the first factor manipulated between subjects. The design for testing was a 4 (condition) X 3 (test time) mixed factorial, with the first factor manipulated between subjects. The first variable was condition. There was a pre-test, four study-test rounds, and an immediate post-test on Day 1. Both the pre-test and the post-test included all 64 facts without the opportunity for prior study. On Day 2 there was a retention test, which was equivalent to the pre-test and the post-test. In the four study-test rounds, the first and last rounds included all 64 facts for all conditions. The facts presented in the second and third rounds varied between conditions. The conditions were the (a) full-study, (b) dropout, (c) yoked, and (d) clicker conditions. There were 12 participants in each of the full-study and dropout conditions, and there were 11 participants in each of the yoked and clicker conditions. For the full-study condition, participants studied and were tested over all of the facts in every round. In the dropout condition, mastered facts were dropped from subsequent rounds. Participants in the yoked condition saw only the facts that a matched participant in the dropout condition saw. Participants in the clicker condition saw the average number of facts presented in Rounds 2 and 3 of the dropout and yoked conditions. The facts selected for presentation in Round 2 were the top 41 missed facts in Round 1 of the fullstudy condition, and the facts presented in Round 3 were the top 28 missed facts in Round 2 of the full-study condition. The remaining variables, learning round (for the analysis of learning) and test time (for the analysis of testing), were manipulated within subjects. The dependent variable was the proportion of correct responses.

As in Experiment 1, the clicker condition is a laboratory analogue of how response clickers are used by instructors in the classroom, such that all of the participants in the clicker

condition were tested over material that was missed by a portion of participants in the full-study condition. Again, facts at test were presented as cued recall rather than multiple-choice in order to minimize the effects of guessing. Responses were hand-scored so that responses that were incorrect due only to misspellings were counted as correct.

#### Materials

Experiment 2 utilized the same fact-learning task as Experiment 1. The same set of 64 true facts was used as in Experiment 1.

#### Procedure

The same procedure was used as in Experiment 1 except for the addition of the three tests. On Day 1, participants were informed that they would be viewing several sets of incomplete facts about countries and that they were to fill in the blanks if they knew the answer. This was the pre-test, which was used to assess the knowledge of participants upon entry to the experiment. Following the pre-test, participants were instructed to read a second set of instructions that informed them that they would be viewing several sets of eight facts about countries and that they would be viewing several sets of eight facts about countries and that they would be viewing several sets of eight facts about countries and that they would be tested on their ability to recall those facts. The experimenter then initiated the presentation of the fact lists, in four study-test rounds, as in Experiment 1. After participants completed all four study-test rounds, they were given a post-test, which was equivalent to the pre-test in that it included all 64 facts. As in the pre-test, facts were randomized within their respective blocks, but the blocks were in the same order as during training.

On Day 2, participants completed a retention test, which was equivalent to both the preand post-tests, such that it also included all 64 facts. As in the pre- and post-tests, the presentation of facts within their respective blocks was random with block order the same.

Participants were given no opportunity to study the facts before completing this retention test. After participants completed the experimental session on Day 2, they were debriefed.

#### Results

A 4 (condition) x 2 (Round 1 vs. Round 4) ANOVA for the analysis of accuracy during learning revealed a main effect of condition, F(3, 42) = 12.40, MSE = 0.26, p < .0001, with accuracy being highest in the full-study condition (M = .51), followed by the dropout (M = .39), the yoked (M = .24), and the clicker (M = .24) conditions. The analysis of accuracy during learning also revealed a marginally significant main effect of learning round, F(1, 42) = 3.84, MSE = 0.10, p = .0568, with accuracy tending to be higher in Round 4 (M = .37) than in Round 1 (M = .33) (see Figure 2). A Fisher's PLSD post-hoc analysis, with an alpha of .05, was conducted for accuracy during learning. The post-hoc analysis showed a significant difference between the full-study condition and the clicker (p < .0001), yoked (p < .0001), and dropout (p = .0288) conditions. This post-hoc analysis also showed a significant difference between the dropout condition and the clicker (p = .0064) and yoked (p = .0055) conditions. Finally, the post-hoc analysis showed no significant difference between the clicker condition and the yoked condition.



*Figure 2*. Mean proportions of correct recall for Block 1 and Block 4 as a function of learning condition in Experiment 2. Error bars represent +/- 1 standard error of the mean.

A 4 (condition) x 3 (test time) ANOVA for the analysis of accuracy at test revealed a main effect of condition, F(3, 42) = 7.60, MSE = 0.14, p = .0004, with accuracy being highest in the full-study condition (M = .21), followed by the dropout (M = .12), the yoked (M = .10), and the clicker (M = .07) conditions. The analysis of accuracy at test also revealed a main effect of test time F(2, 84) = 84.13, MSE = 0.04, p < .0001, with accuracy being lowest at the pre-test (M = .02), highest at the immediate post-test (M = .22), and intermediate at the retention test (M = .14). A Fisher's PLSD post-hoc analysis, with an alpha of .05, was conducted for accuracy during testing. The post-hoc analysis showed a significant difference between the full-study condition and the clicker (p < .0001), the dropout (p = .0042), and the yoked (p = .0007) conditions. There was no significant difference between the clicker or the yoked conditions, and there was also no significant difference between the clicker and yoked conditions.

The analysis of accuracy at test also revealed a significant interaction between condition and test time, F(6, 84) = 6.77, MSE = 0.04, p < .0001 (see Figure 3). At the immediate post-test, accuracy was highest in the full-study condition (M = .37), followed by the dropout condition (M = .22), the yoked condition (M = .17), and then the clicker condition (M = .10). At the retention test 1 week later, accuracy was highest in the full-study condition (M = .24), followed by the dropout condition (M = .13), the yoked condition (M = .11), and then the clicker condition (M = .10), with the clicker condition showing less loss across the 1-week delay than any of the other conditions.



*Figure 3*. Interaction between test time and condition in Experiment 2. Error bars represent +/- 1 standard error of the mean.

Difference scores were computed in order to assess forgetting across the 1-week delay between the immediate post-test and the retention test. The analysis of difference scores revealed a main effect of condition, F(3, 42) = 4.04, MSE = 0.13, p = .0130, with the most forgetting occurring in the full-study condition followed by the dropout, the yoked, and then the clicker conditions (see Figure 4). A Fisher's PLSD test was conducted, with an alpha of .05, which revealed that forgetting in the clicker condition was significantly less than forgetting in the dropout (p = .0226) and the full-study (p = .0017) conditions.



*Figure 4*. Forgetting from the immediate test to the retention test as a function of condition in Experiment 2. Error bars represent +/- 1 standard error of the mean.

#### Discussion

Experiment 2 replicated the results of Experiment 1 and previous research using the dropout procedure (Karpicke & Roediger, 2008; Rock, 1957), in that the dropout condition performed better than the yoked condition, showing an advantage of completely individualized over non-individualized training. However, this difference between the dropout and yoked conditions also occurred in Round 1, where there should have been no difference.

Based on the results of Experiment 1, it was predicted that the clicker condition would perform significantly better than the yoked condition during learning and at test, but Experiment 2 revealed no difference between the clicker and the yoked conditions during learning or at test, suggesting no advantage for group-based learning over non individualized learning. This result would suggest that using the clicker technique, which tailors lecture according to a group's understanding, is no better than a traditional lecture style, which does not tailor lecture to students' understanding. Both the clicker and the yoked conditions performed significantly worse during learning than each of the dropout and the full-study conditions, and both the clicker and the yoked conditions performed significantly worse during testing than the full-study condition. This result suggests that completely individualized learning is more effective in the short-term than group-based or no individualized learning. As would be expected, repeatedly studying and testing all material, as in full-study, is more effective than studying material that another person may need to study and test, as in the yoked condition. Surprisingly, this result also suggests that repeatedly studying and testing over all material is more beneficial than the compression provided by the clicker technique. It is possible that the results from Experiment 1 that did not replicate in Experiment 2 are due to the order in which conditions were tested in the semester. Due to the nature of the experimental design, each participant in the yoked condition had to be tested after his or her matched participant in the dropout conditions.

The dropout condition performed significantly worse during learning than the full-study condition. The latter result was unexpected, suggesting that perhaps it is sometimes beneficial to study and test over all material, but it is likely that as the amount of to-be-learned material increases, the limitations of the full-study procedure and the advantages of the dropout procedure may grow more pronounced. Given that the current set of 64 facts was drawn from a larger fact set of 144 facts, it is possible that 64 facts may have been too few to guarantee enough of an information overload for the full-study condition to be vulnerable to the adverse effects of studying and testing every item every time.

Interestingly, the interaction between test time (pre, post, retention) and condition showed a decrease in performance between the post-test and the retention test for all conditions except the clicker condition. Although the overall performance on the post- and retention tests of the clicker condition was worse than the performance of both the dropout and full-study conditions,

the clicker condition was the only condition that did not show a decrease in performance over the 1-week retention interval. The latter result is important because the knowledge learned by participants in the clicker condition was more durable over time than the knowledge learned by participants in the dropout and full-study conditions. Perhaps the criteria of using group performance to select the facts for presentation in the clicker condition was a more reliable estimate for determining problematic facts than the criteria of using an individual's performance in the dropout condition. The clicker condition's more reliable estimate of problematic facts may have introduced an appropriate level of difficulty during learning to lead to durability of learned facts, while compressing, rather than expanding, the intervals between learning and testing. The pattern of performance of the clicker condition (i.e., low performance during learning coupled with durability) is consistent with research (Schmidt & Bjork, 1992; Schneider et al., 2002) showing that lower performance during learning, due to the introduction of difficulty during learning, produces greater durability. Experiment 3 examined again the durability of knowledge learned under the clicker technique while controlling for the condition order problems in Experiments 1 and 2 and also directly examined the transferability of knowledge learned under the clicker technique.

#### CHAPTER IV

#### **EXPERIMENT 3**

Educational tools are only useful to the extent that they lead to knowledge that will be retained over extended periods of time and that can be applied in novel situations. Experiment 1 showed that a laboratory analogue of the clicker technique facilitates initial fact acquisition while reducing study time; however, neither Experiment 1, nor Experiment 2, nor any previously published study of the clicker technique, has provided any evidence on whether or not the knowledge acquired with the clicker technique is generalizable. Experiment 3 was conducted in order to examine how the learning conditions from Experiment 2 would affect the retention of facts from a novel fact set over a 1-week delay and to examine whether or not that fact knowledge would generalize to a related question both immediately and 1 week later. In order to investigate generalization of knowledge, each fact had two forms, a general form and a specific form. Half of the facts changed forms between learning and testing. Experiment 3 was designed to be more compatible with how clickers are used in the classroom, such that the facts were tested as multiple-choice questions (unlike those in Experiments 1 and 2, which involved cued recall). Additionally, two norming groups were included so that the four experimental conditions could be tested simultaneously in order to eliminate any confounding effects that would result from the nature of the experimental design, which would otherwise force some conditions to be tested later in the semester than other conditions, as in Experiments 1 and 2.

Smith et al. (2009) found that students' performance increased between a clicker question and an isomorphic question. This increase in performance was attributed to the small group discussion period that took place in between these two question sets. Although discussion is known to be a valuable learning tool, the time spent on discussion is exchanged for other

material in the curriculum that needs to be covered. Depending upon how much time is spent on discussions, the content of planned lectures must be changed on the fly, resulting in lost material. Given that clickers are often used in very large introductory level classes, this loss of material may have significant implications for students who intend to pursue higher-level courses in a given subject. Such students may not have all of the base knowledge required for more advanced courses.

The procedural reinstatement principle describes the tendency for declarative information to be rapidly lost over time but highly generalizable (Healy, 2007; Healy & Bourne, 1995). According to the procedural reinstatement principle, facts acquired under the clicker technique should be rapidly forgotten but generalizable because the information learned is declarative. However, we know from the literature on the testing effect (Carpenter & DeLosh, 2005; Karpicke & Roediger, 2008) that testing leads to learning and more specifically, that testing after initial presentation leads to slower forgetting (Carpenter et al., 2008). Given that the clicker technique introduces tests, which promote durability of declarative information, which is normally rapidly forgotten, it might be expected that knowledge acquired with the clicker technique can be both durable and generalizable, while also conserving learning time. Experiment 3 explored the effects of the clicker technique, via its tendency to compress learning time based on group performance, on the acquisition, retention, and generalizability of knowledge on both immediate and delayed tests.

#### Method

#### **Participants**

Forty-seven undergraduate University of Colorado students participated in the experimental conditions, and an additional 24 students participated in the norming conditions in

order to partially fulfill requirements for an introductory psychology course. The initial participants were assigned by a fixed rotation to the two norming conditions, with 12 participants in each norming condition. Subsequent participants were assigned by a fixed rotation to the four experimental conditions, with 12 participants in each condition, except the full-study condition, which contained only 11 participants because of an experimenter error.

#### Design

Prior to conducting the four experimental conditions, two groups of 24 students participated in learning Rounds 1-4 and in an immediate post-test of the 64 facts, without an opportunity for prior study. These groups were called the dropout and full-study norming conditions. It is from these participants' data that the facts that were presented in the clicker condition were based. The clicker condition is a laboratory analogue of how response clickers are used by instructors in the classroom, such that all of the participants in the clicker condition were tested over the same amount of material that was missed by participants in the dropout norming group. Facts at quiz and at test were presented as multiple-choice questions in order to mimic the method of presentation of clicker questions in the classroom.

The experimental conditions were the (a) full-study, (b) dropout, (c) yoked, and (d) clicker conditions. For the full-study condition participants studied and quizzed over all of the facts in every round. In the dropout condition, mastered facts were dropped from subsequent rounds. Participants in the yoked condition saw only the facts that a matched participant in the dropout condition saw. On average, participants in the dropout and yoked conditions saw 20 (of 64 possible) facts in Round 2 and 5 facts in Round 3, representing a 69% and 92% compression in Rounds 2 and 3, respectively. Finally, the clicker condition participants saw the average number of facts presented in Rounds 2 and 3 of the dropout norming condition. The facts

selected for presentation in Round 2 were the top 26 (of 64 possible) missed facts in Round 1 of the full study norming group, and the facts presented in Round 3 were the top 10 missed facts in Round 2 of the full-study norming group, representing a 59% and 84% compression in Rounds 2 and 3, respectively.

The design for learning is a  $4 \ge 2 \ge 2$  mixed factorial. The first factor of condition (fullstudy, dropout, yoked, clicker) was manipulated between subjects. The second factor of learning round (1 vs. 4) and the third factor of learning fact format (general, specific) were both manipulated within subjects.

The design for test is a  $4 \ge 2 \ge 2 \ge 2 \ge 2$  mixed factorial. The first factor of condition (fullstudy, dropout, yoked, clicker) was manipulated between subjects. The second factor of test time (immediate, retention), the third factor of learning fact format (general, specific), and the fourth factor of test fact format (general, specific) were all manipulated within subjects. For both learning and test, the dependent variable examined was accuracy.

#### Materials

The current study utilized a fact-learning task, which consists of 64 facts about eight different plant categories. Each plant category had eight different exemplars whose names were fictitious. All of the fictitious plant names were generated from actual plant names. Twenty to 30 plant names in a given plant category were entered into a word generator, which sliced and diced the entered plant names into novel names, which read like real words. The facts were presented as sentences, each including a plant category, verb phrase, and name. Each fact for a given fake plant is true for a given, matched real plant. Each fact for each plant exemplar was presented in two forms: a general form (e.g., "A tree that comes from Asia is the *Pawthra*") and

a specific form (e.g., "A tree that is native to southern India is the *Pawthra*"). The italicized word (the fake plant name) is what participants were tested on at the multiple-choice quiz.

Each four-option multiple-choice quiz question had two within-set distractors (i.e., from the same plant category) and one out-of-set distractor (i.e., a fictitious plant name generated from plant names of the same plant category that was not included as an exemplar of any of the plant categories). Each fictitious plant name exemplar was used once as a correct answer and twice as a distractor answer, except one exemplar in each category was used three times as a distractor answer. Each fact for all plant categories had a single one-word answer, which was the fictitious plant name (see Appendix B).

#### Procedure

*Norming conditions.* Participants in the norming groups were tested individually in separate rooms on Apple i-Mac computers. Participants were informed that they would be viewing several sets of eight facts about different types of plants and that they would be tested on their ability to recall eight facts about those same plants. The experimenter then initiated the presentation of the fact lists. Each fact was presented individually for 3 s in blocks of eight facts, with all eight exemplars of a given plant category in each block of the full-study norming condition. Within each block of the full-study norming condition, participants studied four general plant facts and four specific plant facts. After each block, participants were given a multiple-choice quiz over the four general and four specific plant facts that they had just studied. Subjects studied a given question the same way on each of the four rounds. That is, during the four learning rounds, the questions that were presented in each block as general and specific were consistent during both study and quiz. The general and specific facts were counterbalanced between the learning phase and the immediate post-test, such that half of the general facts during

learning remained in general format during testing and the other half of the general facts during learning switched to specific format during testing. Likewise, half of the specific facts during learning remained in specific format during testing and the other half of the specific facts during learning switched to general format during testing. Across participants in each condition a given fact occurred in each of the four format combinations at learning and at test. Participants were given the plant name and a verb phrase, followed by a blank (e.g., "A tree that comes from Asia is the \_\_\_\_\_.") and selected the appropriate answer from the four possible multiple-choice alternatives. Participants were given 9-s to begin a response before the program automatically proceeded to the next question.

There were four study-quiz rounds, each consisting of eight blocks. The order of blocks was constant, and the same eight facts occurred within each block, but the order in which the facts were presented was random within the blocks. As in Experiments 1 and 2, the order of the presentation and testing of the facts in a block were different so that participants could not use a serial order mnemonic strategy. In the full-study norming condition, Rounds 2 and 3 included all 64 facts. In the dropout norming condition, Rounds 1 and 4 were like those in the full-study norming condition, but Round 2 consisted of only those facts missed in Round 1, and Round 3 consisted of only those facts missed in Round 2. Upon completion of Rounds 1-4, the two norming groups participated in an immediate post-test, which included all 64 facts, without the opportunity for prior study. In the immediate test, facts were presented within the same blocks as they were during Rounds 1-4, but the order of presentation of facts was again randomized within each block.

*Experimental conditions.* The four experimental conditions participated in Rounds 1-4 and the immediate post-test described above, just as the two norming groups did. Participants in

the experimental conditions also returned 7 days after the initial experimental session to complete a retention test. In the dropout, yoked, and clicker conditions, in which not all of the facts were included in Rounds 2 and 3, facts were still presented randomly within their respective blocks. Both the immediate post-test and the retention test included all 64 facts, without the opportunity for prior study. In both the immediate post-test and retention test, facts were presented within the same blocks as during learning, and facts were randomized within their respective blocks separately for each test so that participants could not use a serial order mnemonic strategy. Participants in the yoked condition were always tested directly after their matched participant from the dropout condition because the facts included for presentation in Rounds 2 and 3 of the yoked condition depended on the performance of the matched subject from the dropout condition. Subjects in the clicker condition received the same number of facts in Rounds 2 and 3 as the average number in the dropout norming condition (26 and 10, respectively). The particular facts shown were the ones missed most often in the full-study norming condition on Rounds 1 and 2, respectively.

It was during the immediate post-test on Week 1 and the retention test on Week 2 that the general and specific question transfer manipulation occurred. Following the completion of the four study-quiz rounds, participants completed a multiple-choice post-test of all 64 facts without opportunity for study. At the immediate post-test, for each question type (specific or general) half of the items were in the same format as at study (i.e., general/general or specific/specific) and half were switched (general/specific or specific/general). The latter transfer manipulation occurred in all of the experimental conditions. The retention test was also a multiple-choice test of all 64 facts without opportunity for study. At the retention test, all questions for a given subject were in the opposite format as on the immediate post-test.

#### Results

#### Norming

An analysis of variance (ANOVA) on accuracy at test was employed with condition (fullstudy norming, dropout norming) as the only factor. The analysis did not reveal a main effect of condition, F(1, 46) < 1.

#### Learning

A 4 (condition) x 2 (Round 1 vs. Round 4) x 2 (learning format) ANOVA on accuracy during learning was employed. The analysis of accuracy revealed only a main effect of learning round, F(1, 43) = 222.46, MSE = 0.02, p < .0001, with accuracy being higher in Round 4 (M =.85) than in Round 1 (M = .62), demonstrating fact learning. A Fishers PLSD test, with an alpha of .05, was conducted, which revealed that performance during learning in the yoked condition (M = .667) was significantly worse than that of the dropout (M = .788, p = .0357) and clicker conditions (M = .778, p = .0424), with the full-study condition's (M = .709) performance in the middle.

#### Test

A 4 (condition) x 2 (test time) x 2 (learning format) x 2 (test format) ANOVA on accuracy at test was employed. Overall, the analysis of accuracy revealed a main effect of learning format, F(1, 43) = 7.89, MSE = 0.02, p = .0074, with accuracy at test being higher when learning occurred with specific facts (M = .62) than with general facts (M = .59). Interestingly, there was a significant interaction between learning format and condition, F(3, 43) = 3.15, MSE= 0.02, p = .0345, with performance in all conditions, except the full-study condition, being higher during testing when subjects learned with specific facts (see Figure 5).



*Figure 5*. Interaction at test between learning format and condition in Experiment 3. Error bars represent +/- 1 standard error of the mean.

The ANOVA also revealed a main effect of test time, F(1, 43) = 178.66, MSE = 0.02, p <.0001, with accuracy being higher at the immediate test (M = .70) than at the retention test (M = .51), demonstrating forgetting across the 1-week retention interval. There was a significant interaction between test time and condition, F(3, 43) = 4.08, MSE = 0.02, p = .0124, indicating that the performance decline from the immediate test to the retention test differed between conditions (see Figure 6). At the immediate test, performance was better for the clicker, the dropout, and the full-study conditions than for the yoked condition. At the retention test, the advantage of the clicker condition was slightly reduced, but overall, a similar pattern was evident as at the immediate test with better performance for the clicker, dropout, and full-study conditions.



*Figure 6*. Interaction at test between test time and condition in Experiment 3. Error bars represent +/- 1 standard error of the mean.

Also, the analysis revealed an interaction between learning format and test format, F(1, 43) = 106.43, MSE = 0.02, p < .0001, with the number of correct responses at test being greater when the learning and testing formats were the same (both specific, M = .70; both general, M = .66) than when they differed (specific/general, M = .55; general/specific, M = .51). There was a significant interaction of test time, learning format, and condition, F(3, 43) = 3.36, MSE = 0.01, p = .0273, with a disadvantage for the clicker condition evident only at the retention test when participants learned with general facts (see Figure 7).



*Figure 7*. Three-way interaction at test between test time, learning fact format, and condition in Experiment 3. Error bars represent +/- 1 standard error of the mean.

Finally, the analysis of accuracy revealed a three-way interaction between test time, learning format, and test format, F(1, 43) = 9.44, MSE = 0.02, p = .0037, showing that the higher performance when learning and testing format matched than when they differed was greater for the immediate test than for the retention test. In addition, there was some evidence of transfer of knowledge from one format to another because performance was well above chance, .25, in all cases even at the retention test (see Figure 8).



*Figure 8*. Three way interaction at test between test time, learning fact format, and test fact format in Experiment 3. Error bars represent +/- 1 standard error of the mean.

Difference scores were computed in order to assess forgetting across the 1-week delay between the immediate test and the retention test. The analysis of difference scores revealed a main effect of condition, F(3, 43) = 4.08, MSE = 0.04, p = .0124, with the most forgetting occurring in the clicker condition followed by the dropout, the full-study, and then the yoked conditions (see Figure 9). A Fisher's PLSD test was conducted, with an alpha of .05, which revealed that forgetting in the clicker condition was significantly greater than forgetting in the dropout (p = .0245), full-study (p = .0177), and yoked (p = .0018) conditions.



*Figure 9*. Forgetting between the immediate test and the retention test as a function of condition in Experiment 3. Error bars represent +/- 1 standard error of the mean.

An ANOVA on accuracy, restricted to the rention test, was employed. The analysis did not reveal a main effect of condition, F(3, 43) = 1.21, MSE = 0.11, p = .3194, demonstrating no differences in final performance among conditions.

A 4 (condition) x 2 (test time) x 2 (learning format) x 2 (test format) x 4 (item difficulty) ANOVA on accuracy at test was conducted in order examine the role of item difficulty, with items being classified as having an easy, easy-medium, medium-hard, or hard difficulty level. Only the results involving the factor of difficulty are reported here. Facts were classified into difficulty levels on the basis of the average performance of participants in the full-study norming condition on all 64 facts across all four learning rounds. The ANOVA revealed a main effect of item difficulty, F(3, 129) = 18.54, MSE = .05, p < .0001. A Fischer's PLSD, with an alpha of .05, showed that that accuracy highest for easy facts (M = .67), followed by easy-medium facts (M = .63), followed by medium-hard facts (M = .57) and hard facts (M = .56). Approaching significance was an interaction between test time and item difficulty, F(3, 129) = 2.16, MSE = .06, p = .0955, with accuracy at the immediate test greatest for easy facts, followed by easymedium facts, followed by medium-hard facts, and lowest for hard facts and accuracy at the retention test revealing a similar pattern except accuracy for hard facts was higher than accuracy for medium-hard facts (see Figure 10).



*Figure 10.* Interaction between test time and item difficulty in Experiment 3. Error bars represent +/- 1 standard error of the mean.

Finally, the ANOVA revealed an interaction between item difficulty, learning format, and test format, F(3, 129) = 5.00, MSE = .04, p = .0026, demonstrating more item format specificity for easier facts than for harder facts and, interestingly, more transfer from specific to general



facts than from general to specific facts only for facts with a hard difficulty level (see Figure 11).

*Figure 11*. Three-way interaction between item difficulty, learning fact format, and test fact format in Experiment 3. Error bars represent +/- 1 standard error of the mean.

#### Discussion

Experiment 3 was conducted to examine how knowledge acquired with the clicker technique compared to knowledge acquired with other learning conditions in terms of its durability and generalizability over a 1-week retention interval. In this experiment participants learned unfamiliar facts (in either a general or specific format) about plants, and the learning phase either involved no compression, compression based on the performance of the tested participant, compression based on the performance of another participant, or compression based on the performance of a group of participants. Participants in all conditions demonstrated fact learning from the beginning to the end of the learning phase.

During the testing phase of the experiment, half of the facts switched format from the learning phase, and it was shown that performance during testing was higher when learning occurred with specific facts than with general facts for all conditions, except the full-study condition. Previous research has shown that greater elaborative encoding (i.e., greater breadth or amount of processing) increases the likelihood that a distinctive feature will be encoded (Eysenck & Eysenck, 1979; Winograd, 1981). One possible explanation for the advantage of studying with specific facts is that the specific forms of facts contained more distinguishing and meaningful information than the general forms of facts. The amount of information in the specific facts may have promoted more elaborative processing during learning, increasing the encoding of distinctive features. All conditions, except the full-study condition, which was the only condition that did not benefit from learning specific facts, contained some sort of learning time compression. Learning time was reduced in the clicker condition by a total of 59% in Round 2 and 84% in Round 3, and learning time was reduced in the dropout and yoked conditions by a total of 69% in Round 2 and 92% in Round 3. The compression in these conditions might have freed up processing capacity, which was more limited in the full-study condition, thus allowing a more detailed level of learning, as might be offered by the specific fact format.

As expected, forgetting between the immediate post-test and the retention test occurred for all conditions, which is consistent with the procedural reinstatement principle, which describes the tendency for declarative information to be rapidly lost over time (Healy & Bourne, 1995). The clicker, dropout, and full-study conditions all performed better than the yoked condition at both the immediate post-test and the retention test, demonstrating that compression based on the performance of the individual (dropout condition), compression based on the performance of a group of people (clicker condition), and no compression (full-study) during learning is better than compression based on the performance of another individual (yoked condition). Although the advantage of the clicker condition was slightly reduced at the retention

test, the pattern of results was similar to that at the immediate post-test (see Figure 5). Contrary to the results of Experiment 2, the analysis of difference scores revealed that the clicker condition forgot more information than all of the other conditions between the immediate and retention tests, suggesting that knowledge acquired under the clicker technique is less durable. However, given the results of Experiment 2, demonstrating that knowledge acquired under the clicker technique was most durable, and given the finding that in Experiment 3 the clicker condition's final performance (on the retention test) was comparable to the final performance of the other conditions, the issue of durability of knowledge acquired under the clicker technique is unresolved and needs further investigation.

Performance was greatest during testing when the fact format matched between learning and testing. This result is not surprising because in these conditions, the acquisition activity was identical to the testing activity, which, according to transfer appropriate processing (Morris, Bransford, & Franks, 1977), facilitates the retrieval of memory traces. Interestingly, the clicker condition showed a disadvantage at the retention test only when participants learned general facts during the learning phase. With compression based on the average performance of a larger group, it is more beneficial (for durability) to learn specific facts, but with no compression there was no advantage for learning specific over general facts. The compression used in the clicker technique might, therefore, be a more efficient method for isolating facts that are, on average, more difficult for most people, thus resulting in a greater learning advantage when learning time is spent on these more difficult facts, but only if they are in the more specific format that allows for the encoding of distinctive features.

On the immediate test, performance was lowest on facts with a hard difficulty level, but at the retention test, performance on hard facts was greater than performance on medium-hard

facts. This result is consistent with previous research demonstrating that the introduction of more effortful processes during learning leads to low performance during learning but improved performance on delayed tests (Healy & Bourne, 1995; Schmidt & Bjork, 1992; Schneider, Healy, & Bourne, 2002).

A secondary purpose of Experiment 3 was to explore the generalizability of knowledge acquired with the clicker technique. Although performance was higher during both tests when learning and testing format matched, performance when learning and testing format differed was well above chance level during both tests, demonstrating generalization of knowledge. This result is consistent with the procedural reinstatement principle because the declarative information that was retained over the retention interval was indeed generalizable (Healy, 2007). The latter result is important because it demonstrates that learning with the clicker technique can indeed promote learning of knowledge that is generalizable. Interestingly, the specificity advantage found at test was more pronounced for easier facts than for harder facts and generalization from specific to general fact formats was greater than generalization from general to specific fact formats for facts with a hard difficulty level.

#### CHAPTER V

#### GENERAL DISCUSSION

Learning. The first objective of this study was to examine how learning under a laboratory analogue of the clicker technique compared to learning under other repeated testing schedules. Previous research (Pyc & Rawson, 2007; Rock, 1957) has shown that the dropout procedure is an effective method of compressing individual study time by dropping known items from future studying and testing. Final performance under the dropout procedure tends to be as good as performance under full-study conditions (Pyc & Rawson, 2007). The present study demonstrates that the clicker technique is an effective method of compressing teaching time by dropping items known by the majority of a group from future teaching, without any sacrifice to amount learned. Overall, the clicker conditions (Clicker 25 and Clicker D in Experiment 1 and clicker in Experiment 3) performed just as well as the dropout and full-study conditions and performed better than the yoked condition during learning. This pattern of results demonstrates that the performance level of a group is just as useful as the performance of a given individual and more useful than the performance of another single individual for promoting learning during an acquisition phase. Taken together, these results demonstrate that the clicker technique is effective at promoting group learning, in which the performance of every individual cannot be addressed, that is more effective than teaching that does not consider the understanding of multiple individuals in a group.

*Durability.* The second objective of this study was to examine the extent to which the clicker technique promotes the learning of durable knowledge. Previous research has shown that declarative information is rapidly lost over time (Healy & Bourne, 1995), that greater difficulty of acquisition improves retention (Schneider et al., 2002), that shorter lists are remembered better

than longer lists (Gronlund & Elam, 1994; Strong, 1912), that testing enhances retention (Roediger & Karpicke, 2006), and that more spacing between testing events promotes retention (Hogan & Kintsch, 1971; Melton, 1967; Schmidt & Bjork, 1992). Given that declarative information is rapidly lost over time, the declines in performance observed across the 1-week delay in Experiments 2 and 3 were expected. Although the results of Experiment 2 were likely affected by methodological and confounding circumstances, the lack of loss of information in the clicker condition motivated further exploration of durability in Experiment 3, which controlled for these confounding variables.

Experiment 3 demonstrated that knowledge acquired under the clicker technique was the least durable over a 1-week delay; however, the clicker condition's level of performance at the retention test was still comparable to that of the dropout and full-study conditions. The clicker, dropout, and yoked conditions all involved compression between testing events, via a reduction in the number of tested items, and essentially differed only in the difficulty of the items presented in compressed learning rounds. The spacing effect would predict that the reduced space between testing events in these conditions should hurt retention performance relative to the full-study condition. Conversely, the list-length effect would predict that the reduction in the number of to-be-tested items across learning rounds in these conditions should aid remembering relative to a full-study condition.

Interestingly, the lower overall test performance of the yoked condition, relative to the clicker and dropout conditions, indicates that a reduction in the number of tested items across learning is not enough to aid remembering. Given that these conditions differed only in the criteria that determined the particular items presented, the testing of particular items may be important for remembering. The higher performance on the retention test of the clicker and

dropout conditions indicates that reducing lists of to-be-tested items to the most difficult items (as determined either by the particular individual or a group of individuals) enhances retention test performance, despite a reduction in the spacing between testing events. This result suggests that more spacing between testing events is not always better than less spacing between testing events. In general, testing of fewer but more difficult items promotes higher retention test performance, in spite of reduced spacing between testing events.

Because the dropout and clicker conditions did not perform differently from one another, no strong claims can be made about benefits of using the performance of a group versus the performance of an individual to determine items for further learning. It can be argued, however, that the performance of a group is a more reliable indicator of difficult items than the performance of one individual because, although the performance of the dropout condition was just as good as that in the clicker condition, the performance of the yoked condition was worse than that of the clicker condition, suggesting that the performance of one individual is not always a useful indication of material that needs further study. Furthermore, the facts learned in Experiment 3 varied in difficulty level. When a set of to-be-learned items vary in difficulty, group-performance may be a better indication than self-performance of items for further study and when a set of to-be-learned items are of equal difficulty, self-performance may be a better indication than group-performance of items for further study. In both cases, however, the performance of another single individual is a poor indication of items for further study.

The current study also replicated previous results showing that the dropout procedure can be used as a method of conserving student study time (Pyc & Rawson, 2007). Previous work on the dropout procedure primarily examined associations between single words (e.g., Pyc & Rawson, 2007; Rock, 1957). The current study extended previous results by demonstrating that

the classic results supporting one-trial learning of word pair associations hold in situations that involve learning more complex word-phrase pair associations.

*Generalization.* The third objective of this study was to examine the extent to which the clicker technique promotes the learning of generalizable knowledge. Previous research has shown that procedural information tends to be durable and less generalizable and that declarative information tends to be retained poorly, but declarative information that is retained is indeed more generalizable (Healy, 2007; Healy & Bourne, 1995). Experiment 2 (and to some extent Experiment 3) demonstrated that declarative knowledge learned under the clicker technique is retained over time, suggesting that this knowledge may also be generalizable.

Generalization was examined by testing performance when facts either stayed in the same or switched format between learning and testing. The higher test performance when learning and testing format matched than when learning and testing format differed is consistent with transfer appropriate processing theory (Morris et al., 1977), such that retention was better when learning and testing formats were the same. It should be noted that previous studies (e.g., McDaniel, Friedman, & Bourne, 1978) of transfer appropriate processing focus on the consistency between the kinds of *processes* involved in learning and testing and the kind of information being learned, whereas the current study examines consistency between learning and testing of actual *item* formats (as in the study by Roediger & Blaxton, 1987), in the present case differing in level of abstraction.

Interestingly, there was an advantage during testing when participants learned specific facts. Previous research on sentence comprehension shows that after a sentence has been interpreted, the original form of the sentence is not important for remembering the meaning of that sentence (Sachs, 1967). Semantic abstractions from a specific sentence can be applied to the

facts used in this study, such that the two versions of each fact differ in their level of abstraction, with facts in general formats being abstract versions of facts in specific formats. Previous research shows that once the semantic meanings of sentences have been stored, the original forms (i.e., the exact words) of sentences are not important for remembering, but in this study the original format of facts was important because performance was higher when learning occurred with specific facts than with general facts. This result suggests that the original form of a fact is important for retaining meaning, perhaps with specific facts, which are less abstract, increasing the likelihood of encoding a distinctive feature that will help remembering (Eysenck & Eysenck, 1979; Winograd, 1981).

Importantly, for all conditions, generalization between fact formats occurred in both directions (general to specific and specific to general) at the immediate and retention tests, indicating that knowledge that was durable across the 1-week delay was also generalizable. This result is consistent with the procedural reinstatement principle (Healy, 2007), because we found generalizability of declarative knowledge about particular items rather than procedures involved in skill learning. The results indicate that learning of flexible knowledge occurs under the clicker technique, while compressing study time as determined by the performance of a group, without sacrifice of amount learned or amount of generalizability.

*Implications and future direction.* Overall, the current study demonstrates that repeated testing of items determined as difficult by a group, as in the clicker technique, promotes the learning of generalizable and potentially durable knowledge. The advantage for using a group instead of a single individual (other than the individual participant) was shown by the advantage for the clicker condition over the yoked condition. Although no direct learning benefits of using group- over self-performance to determine material for further study can be concluded from this

study, there are important practical benefits. In large class settings it is unrealistic and would be time consuming to cater lecture time according to every individual in a class. Given that using group-performance to tailor lecture time produces learning and retention that is just as good as using individual self-performance, instructors do not need to spend extra lecture time on material that the majority of the class understands. The results of these experiments demonstrate a simple and efficient method, using testing and the performance level of a group, to determine material to be included in or dropped from further teaching time, that can promote successful learning in situations that make it difficult for instructors to attend to the individual learning needs of a large number of people. Testing via clicker questions can efficiently help instructors determine which material would be most useful to cover more elaborately during lecture time in order to maximize learning benefits for the majority of students.

In the current study, the laboratory analogue of the clicker technique reduced learning time (e.g., in Experiment 3 by 59% in Round 2 and by 84% in Round 3) but the conserved time remained unused. In the classroom, this extra time can be can be devoted to in-class discussions, as proposed by Smith et al. (2009) and Duncan (2005), or it can be devoted to further elaboration of troublesome material. Future work should include an exploration of various uses of the time conserved under the clicker technique. One potential use of the conserved time is to study and test over both general and specific forms of the facts. Two additional potential uses of the conserved time might be either for additional studying and testing of missed items or for additional studying and testing of rephrased versions of missed items, so that overall learning time is equated across learning conditions. The additional time spent on missed facts might improve performance in the clicker and dropout conditions above the full-study condition. The latter

would provide further support for group performance being a more reliable indicator of largely misunderstood material than individual (self or other) performance.

The testing advantage of specific over general facts observed in Experiment 3 can be further explored in relationship to the recently proposed mediator effectiveness hypothesis (Pyc & Rawson, 2010). According to this hypothesis, testing (versus restudying) facilitates the use of more effective mediators (e.g., a concept or word linking a cue and target), which are considered more effective when the mediators themselves can be retrieved and when the mediators lead to target retrieval. Mediators were participant-generated words that looked or sounded similar to the cue and were semantically related to the target, but were not included in either the cue or the target (Pyc & Rawson, 2010).

In the current study's Experiment 3, the advantage for specific over general facts was explained by suggesting that specific facts may have more distinctive features than general facts and that repeated testing increased the likelihood that one or more of these distinctive features were encoded. Perhaps these distinctive features serve as a type of mediator that is present within the cue itself. In order to determine if there are distinctive words in specific and general facts that serve as mediators, future work should examine participants' abilities to recall words contained in cue phrases when presented with a target and also examine the relationship between general vs. specific cue words recalled and target recall. Once potential cue mediators have been identified, it would be interesting to test target recall following the presentation of only cue mediators. In the current study, the targets (i.e., plant names) were not semantically related to the cues (i.e., general or specific verb phrase) because the target names were fictional, suggesting that a semantic relationship to the target may not be a necessary property of effective mediators. A mediator may more generally be something that is added during learning that makes a

relationship between a cue and a target more distinctive, whether it is independent of the cue and target (as in Pyc & Rawson, 2010) or added to the actual cue itself (as in the current Experiment 3 with general and specific facts).

#### References

- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185-205). Cambridge, MA: MIT Press.
- Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures? *Applied Cognitive Psychology*, 23, 747-759.
- Carpenter, S. K., & Delosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, 19, 619-636.
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36*, 438-448.
- Donovan, W. (2008). An electronic response system and concept tests in general chemistry courses. *The Journal of Computers in Mathematics and Science Teaching*, 27(4), 369-389.
- Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, 20, 81-94.
- Dreifus, C. (2005, November 1). Physics laureate hopes to help students over the science blahs: A conversation with Carl Wieman. *New York Times*, Science Section.
- Duncan, D. (2005). Clickers in the classroom: How to enhance science teaching using classroom response system. San Francisco: Pearson Addison Wesley.
- Eysenck, M. W., & Eysenck, M. C. (1979). Processing depth, elaboration of encoding, memory stores, and expended processing capacity. *Journal of Experimental Psychology*, 5, 472-484.
- Greer, L., & Heaney, P. J. (2004). Real-time analysis of students comprehension: an assessment of electronic students response technology in an introductory earth science course. *Journal of Geoscience Education, 52,* 345-351.
- Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1355-1369.
- Healy, A. F. (2007). Transfer: specificity and generality. In H. L. Roediger, III, Y. Dudai,
  & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 271-275). New York: Oxford University Press.
- Healy, A. F., & Bourne, L. E., Jr. (1995). Preface. *Learning and memory of knowledge and skills: Durability and specificity.* Thousand Oaks, CA: Sage.

- Hogan, R. M, & Kintsch, W. (1971). Differential effects of study and test trials on longterm recognition and recall. *Journal of Verbal Learning and Behavior*, 10, 562-567.
- Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal* of Experimental Psychology, 33, 704-719.
- Karpicke, J. D., & Roediger, H. L., III (2008). The critical importance of retrieval for learning. *Science*, 319, 966-968.
- Karpicke, J. D., & Roediger, H. L., III. (2010). Is expanding retrieval a superior method for learning text materials? *Memory and Cognition*, 38, 116-124.
- Keller, F. S. (1974). Ten years of personalized instruction. Teaching of Psychology, 1, 4-9.
- Kennedy, G. E. & Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning*, 21, 260-268.
- Kole, J. A., & Healy, A. F. (2007). Using prior knowledge to minimize interference when learning large amounts of information. *Memory & Cognition, 35*, 124-137.
- Kulik, J., Kulik, C. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist*, *34*, 307-318.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morries, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.
- Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2008). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, 34, 51-57.
- McDaniel, M. A., Friedman, A., & Bourne, L. E., Jr. (1978). Remembering the levels of information in words. *Memory & Cognition, 6*, 156-164.
- Melton, A. W. (1967). Repetition and retrieval from memory. Science, 158, 532.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal, 16*, 519-533.
- Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, 35, 1917-1927.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator

effectiveness hypothesis. Science, 330-335.

- Rock, I. (1957). The role of repetition in associative learning. *American Journal of Psychology*, *70*, 186-193.
- Roediger, H. L. III., & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory & Cognition*, 15, 379-388.
- Roediger, H. L. III., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science*, *1*, 181-210.
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*, 437-442.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207-217.
- Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign language vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*, 419-440.
- Smith, M. K., Wood, W. B., Adams, W. K., Weiman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323, 122-124.
- Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review*, *19*, 447-462.
- Stuart, S. A. J., Brown, M. I., & Draper, S. W. (2004). Using an electronic voting system in logic lectures: One practioner's application. *Journal of Computer Assisted Learning*, 20, 95-102.
- Trees, A. R., & Jackson, M. H. (2007). The learning environment in clicker classrooms: students processes of learning and involvement in large university-level courses using students response systems. *Learning, Media, and Technology, 32*, 21-40.
- Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology*, 7, 181-190.
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24,* 345-376.

## Appendix A

## Facts used in Experiments 1 and 2 Organized by Country and Fact Category

Country	Verb Phrase
Ghana	major agricultural product is
Niger	currency is the
Rwanda	capital city is
Madagascar	climate is
Malawi	citizens speak
Botswana	major industry is
Solomon Islands	exports goods to
Bahrain	principal religion is

### Fact Categories

Agricultural Product	Capital City	Climate Type	Export Partner
cocoa	Accra	tropical	Netherlands
cowpeas	Niamey	desert	France
coffee	Kigali	temperate	China
vanilla	Antananarivo	variable	U.S.A
tobacco	Lilongwe	subtropical	South Africa
livestock	Gaborone	semiarid	Europe
coconuts	Honiara	monsoon	Korea
fruit	Manama	arid	Saudi Arabia
Languaga Spakan	Major Industry	Official Currency	Principal Religion
Language Spoken	Major muusu y	Official Currency	<u>i i incipal Kengion</u>
English	lumber	Cedi	Pentecostal
English French	lumber mining	Cedi CFA franc	Pentecostal Islam
English French Kinyarwanda	lumber mining cement	Cedi CFA franc R franc	Pentecostal Islam Catholicism
English French Kinyarwanda Malagasy	lumber mining cement meat	Cedi CFA franc R franc MGA	Pentecostal Islam Catholicism Indigenous
English French Kinyarwanda Malagasy Chichewa	lumber mining cement meat tobacco	Cedi CFA franc R franc MGA Kwacha	Pentecostal Islam Catholicism Indigenous Protestantism
English French Kinyarwanda Malagasy Chichewa Setswana	lumber mining cement meat tobacco diamonds	Cedi CFA franc R franc MGA Kwacha Pula	Pentecostal Islam Catholicism Indigenous Protestantism Christianity
English French Kinyarwanda Malagasy Chichewa Setswana Pidgin	lumber mining cement meat tobacco diamonds fishing	Cedi CFA franc R franc MGA Kwacha Pula SI Dollar	Pentecostal Islam Catholicism Indigenous Protestantism Christianity Anglicanism

## Appendix B

Plant Type	General Questions	Specific Questions
Trees	1. A tree that comes from Asia is the <u>Pawthra</u> .	1. A tree that is native to southern India is the <u>Pawthra</u> .
IICC5	2. A tree that is popular in eastern religion is the	2. A tree that is sacred to Hinduism is the <u>Buttony</u> .
	Buttony.	3. A tree that is used for thickening soup is the
	3. A tree that is used in cooking is the <u>Mugwood</u> .	Mugwood.
	4. A tree that is used for athletic equipment is the	4. A tree that is used for making baseball bats is the
	Henbur.	Henbur.
	5. A tree that grows very old is the <u>Boapwort</u> .	5. A tree that can often exceed 3,000 years of age is the
		Boapwort.
	6. A tree that is closely related to flowers is the	6. A tree that is a member of the rose family is the
	Bandpaw.	Bandpaw.
	7. A tree that has unique leaves is the <u>Hawthra</u> .	7. A tree that is identified by its star-shaped leaves is
		the <u>Hawthra</u> .
	8. A tree that is vulnerable to brightness is the	8. A tree that is particularly sensitive to ultraviolet light
	Speetony.	is the <u>Speetony.</u>
Herhs	1. An herb that tastes like a candy is <u>Papwort</u> .	1. An herb that has a licorice flavor is <u>Papwort</u> .
iici b5	2. An herb that is used as a skin treatment is <u>Flace</u> .	2. An herb that is used as a calming remedy for a rash is
	3. An herb that can be used instead of a popular	Flace.
	seasoning is <u>Soabab</u> .	3. An herb that can be used as a salt substitute is
	4. An herb that is used for creating an alcoholic drink is	<u>Soabab</u> .
	<u>Speetrea</u> .	
	5. An herb that is fragile is <u>Clewill</u> .	4. An herb that was one of the first used to brew beers
	6. An herb that has soft leaves is <u>Tandpa</u> .	is <u>Speetrea</u> .
	7. An herb that has colored leaves is <u>Sanyan</u> .	5. An herb that deteriorates if cooked quickly is <u>Clewill</u> .
	8. An herb that is used in a beverage is <u>Boapap</u> .	6. An herb that has very silky leaves is <u>Tandpa</u> .
		7. An herb that grows purple leaves is <u>Sanyan</u> .
		8. An herb that is frequently used for tea is <u>Boapap</u> .
Vines	<ol> <li>A vine that draws insects is the <u>Sper</u>.</li> </ol>	1. A vine that attracts moths is the <u>Sper</u> .
	2. A vine that comes in many varieties is the <u>Silverlat</u> .	2. A vine that has over 400 types and colors is the
	3. A vine that is from South America is the <u>Speedwell</u> .	<u>Silverlat</u> .
	<ol><li>A vine that survives in cold weather is the <u>Swellia</u>.</li></ol>	3. A vine that originated in Brazil is the <u>Speedwell</u> .
		4. A vine that survives in temperatures as low as 20
	5. A vine that invades other plants is the <u>Flatis</u> .	degrees Fahrenheit is the <u>Swellia</u> .
	6. A vine that has leaves like an organ is the <u>Boxwot</u> .	5. A vine that kills off existing vegetation is the <u>Flatis</u> .
	7. A vine that smells good is the <u>Chort</u> .	6. A vine that has heart shaped leaves is the <u>Boxwot</u> .
	8. A vine that helps in first aid is the <u>Bansy</u> .	7. A vine that is known for its sweet fragrance is the
		<u>Chort</u> .
	1 A more all that is makened in the Connell	8. A vine that is used as an antibacterial is the <u>Bansy</u> .
weeds	1. A weed that is robust is the <u>camen</u> .	1. A weed that fives in the cool season is the <u>camen</u> .
	2. A weed that is not found in the west is the Hauman	2. A weed that is found east of the Posleige is the
	A A wood that has a blocky support system is the Sill	S. A weed that is found east of the Rockies is the
	5. A weed that prefers light is the Fanhoo	<u>Hawpaw</u> . A. A weed that has weak square stems is the Sill
	6. A wood that has doon colors is the Soaper	5. A wood that thrives in barsh sun is the Fanboo
	7 A weed that produces seeds in the summer is the	6. A weed that produces blue violet flowers is the
	Flamellia	Sooper
	<u>riancina</u> .	7 A weed that produces seeds beginning in July is the
	8 A weed that has seeds that are dangerous to pets is	Flamellia
	the Timog.	8. A weed that has seeds that can be life-threatening to
	<u></u>	dogs is the Timog.
Wildflowore	1. A wildflower that blooms after the first year is the	1. A wildflower that blooms first on the second year is
winunowei S	Whicory.	the Whicory.
	<u></u>	2. A wildflower that opens at night is the Callyhock.

	<ol> <li>A wildflower that likes darkness is the <u>Callyhock.</u></li> <li>A wildflower that is used for beauty is the <u>Shasty.</u></li> <li>A wildflower that is used as a preventative is the <u>Asta</u>.</li> </ol>
	5. A wildflower that is a kind of vegetable is the <u>Siberie.</u>
	6. A wildflower that is poisonous is the <u>Buttay</u> . 7. A wildflower that is very adaptable is the <u>Tishler.</u>
Fungi	<ul> <li>8. A wildflower that needs a bird to spread its seed is the <u>Bibern</u>.</li> <li>1. A fungus that changes color when it is hurt is the <u>Stinger</u>.</li> </ul>
Shrubs	<ol> <li>A fungus that is shaped like a body part is the <u>Kinkhorn.</u></li> <li>A fungus that is benign is the <u>Inkhort.</u></li> <li>A fungus that defends itself is the <u>Bottine.</u></li> <li>A fungus that defends itself is the <u>Horain.</u></li> <li>A fungus that looks like bone is the <u>Mushen.</u></li> <li>A fungus that doesn't easily dry out is the <u>Direds.</u></li> <li>A fungus that is slick is the <u>Oysted.</u></li> <li>A shrub that is tightly packed is the <u>Ebonbur.</u></li> </ol>
Vegetables	<ol> <li>A shrub that develops in harsh ground is the <u>Henbush</u>.</li> <li>A shrub that draws in insects is the <u>Crasteria</u>.</li> <li>A shrub that withstands bad weather is the <u>Chocolame</u>.</li> <li>A shrub that comes from a distant continent is the <u>Horb</u>.</li> <li>A shrub that is spiky is the <u>Betgum</u>.</li> <li>A shrub that is eaten by mountain wildlife is the <u>Chass</u>.</li> <li>A shrub that develops in moisture is the <u>Sansy</u>.</li> <li>A vegetable that is particularly nutritious is the <u>Caber</u>.</li> </ol>
	<ul> <li>3. A vegetable that has a hot flavor is the <u>Neeper.</u></li> <li>4. A vegetable that can be decorative is the <u>Boreek.</u></li> <li>5. A vegetable that comes in many versions is the <u>Wato.</u></li> <li>6. A vegetable that is popular in Italian cooking is the <u>Kalloof.</u></li> <li>7. A vegetable that irritates the even is the <u>Padive</u></li> </ul>
	7. A vegetable that irritates the eyes is the <u>Kadive.</u>

8. A vegetable that is used as a substitute in cooking is the <u>Cuccoli</u>.

3. A wildflower that is used in make-up products is the
4. A wildflower that has been utilized for birth control is
the <u>Asta</u> . 5. A wildflower that has roots that are wild carrots is
the <u>Siberie</u> .
6. A wildflower that is toxic to the heart is the <u>Buttay</u> .
is the Tishler.
8. A wildflower that depends on Hummingbirds for
pollination is the <u>Bibern</u> .
1. A fungus that turns pink when it is damaged is the
Stinger.
2. A fungus that looks like an ear is the <u>Kinkhorn</u> .
4. A fungus that can close itself off to prodators is the
Rottine
5. A fungus that is often mistaken for dust is the Horain.
6. A fungus that resembles a skull is the <u>Mushen.</u>
7. A fungus that is resistant to dehydration is the <u>Direds</u> .
8. A fungus that appears slimy is the <u>Oysted.</u>
1. A shrub that is known for its compact growth is the
<u>Ebonbur</u> .
2. A shrub that attracts butterflies is the Crasteria
4. A shrub that can survive in harsh winds is the
Chocolame.
5. A shrub that comes from New Zealand is the <u>Horb</u> .
6. A shrub that has needle-like leaves is the <u>Betgum</u> .
7. A shrub that is consumed by bighorn sheep is the
<u>Chass</u> .
8. A shrub that grows in shallow standing water is the
<u>54115y</u> .
1. A vegetable that is in a pie is the <u>Rutaby.</u>
2. A vegetable that is known for its variety of vitamins is
the <u>Caber.</u>
3. A vegetable that has a peppery taste is the <u>Neeper.</u>
4. A vegetable that is often seen as works of art is the
BORCEK.
5. A vegetable that is the foundation of snashotti sauce
is the Kalloof
7. A vegetable that is tear producing when cut is the
Radive.

8. A vegetable that can be used to replace coffee in food preparation is the <u>Cuccoli</u>.