The Pennsylvania State University

The Graduate School

College of Education

EFFECTS OF CONTEXTUAL COLOR ON RECALL: BORDER COLOR AS A LESSON AND POSTTEST CUE FOR FACTUAL AND CONCEPTUAL INFORMATION PRESENTED IN COMPUTER-BASED INSTRUCTION

A Thesis in

Instructional Systems

by

Gustavo E. Prestera

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2003

The thesis of Gustavo E. Prestera has been reviewed and approved* by the following:

Date of Signature

Barbara Grabowski
Associate Professor of Education
Chair of Committee

Roy Clariana
Assistant Professor of Education
Thesis Adviser

Francis Dwyer
Professor of Education

Michael McNeese
Associate Professor of Information Science and Technology

Kyle Peck
Professor of Education
Head of the Department of Adult Education,
Instructional Systems, and Workforce Education

*Signatures are on file in the Graduate School

Abstract

This investigation examined the effects of contextual border colors on recall of the to-be-remembered information, when border colors were displayed on-screen during a computer-based instruction module. The encoding specificity principle (Tulving & Thomson, 1973) posits that contextual cues can only be useful during the posttest if they have first been encoded during the lesson. It was hypothesized that color (a context cue) would enhance recall when present in both the lesson *and* the posttest. Conversely, it was predicted that color would inhibit recall when the color hues in the posttest were intentionally mismatched with those in the lesson. It was also predicted that color effects would differ by gender, item type, and item difficulty. To examine the effects of border colors, a posttest-only control group design was used, with lesson condition and posttest condition as the crossed independent variables, with gender as an additional between-groups independent variable, and with two repeated measures: item type and item difficulty. The dependent variable was cued recall measured with fill-in-the-blank posttest items. An additional outside control group was also used to control for color motivational effects.

Participants were 196 undergraduate students randomly assigned to five treatment groups. Participants read four passages, completed a posttest, read another four passages, and completed a second posttest. The order of reading passages was crossed to offset sequence effects. Depending on the treatment condition, participants saw color-coded border colors in both the lesson and the posttest, no border colors in either, border colors in the lesson and not the posttest, and vice-versa. A fifth group, an additional outside control group, saw border color hues in the posttest that were intentionally mismatched with those seen in the reading passages.

The results do not support the hypothesis that color needs to be encoded during the lesson in order to enhance retrieval, contrary to encoding specificity. The presence of border colors in the *lesson* had positive, non-significant effects on recall of factual and conceptual knowledge, as measured by a difficult fill-in-the-blank posttest. However, the presence of border colors in the *posttest* varied significantly by item type, item difficulty, and gender. Recall of easy factual knowledge cued with verbatim posttest items was significantly higher when border colors were present in the posttest, whether or not border colors were present in the lesson. For males, the presence of border colors in the posttest significantly enhanced recall of the to-be-remembered information. For females, the presence of border colors in the posttest inhibited recall, though this negative effect did not consistently reach significance. These effects are discussed in relation to gender, color, HCI, and context-dependency research.

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgments

Thanks to two wonderful parents, Eduardo and Nilda, I grew up in blissful ignorance of my limitations, knowing only that hard work and pride ensured that even if I did not always reach my destination, I would enjoy the journey. Thank you, mom and dad, you were right!

Along the way, I have been blessed with support beyond measure from friends, colleagues, and mentors. I would like to thank my esteemed committee members for their help, especially, my chair and faculty advisor, Dr. Barbara Grabowski, and my thesis adviser, Dr. Roy Clariana. Roy has guided me every step of the way. I would also like to express my heartfelt appreciation to my friend Dr. Andrew Peck, from the Psychology Department, for his wise counsel and his assistance in preparing the treatment materials and collecting the data for this dissertation. I also appreciate the logistical assistance provided by Nancy Crowder, Nicole Morrone, and the eighteen colleagues who generously volunteered as proctors. I was, and still am, overwhelmed by the outpouring of support from Penn State students, faculty, and staff.

I would not have arrived at this particular junction without the encouragement of some very special people. I would like to thank the late Mike Harfman, who first introduced me to the art and science of instructional design, setting me on my current course. Similarly, Dr. Robert Linneman, friend and mentor since my undergraduate days, has been a source of advice, encouragement, and inspiration throughout my journey. To Dr. Les Moller, thank you for the active support in getting me published early and for your valuable advice. To all of my family, friends, colleagues, and mentors, *mil gracias*, many thanks!

Dedicated to…

My darling wife, Mae, and our young prince, Edward (E.J.).

Your love and support have made all things possible

and humble all attempts to show my appreciation.

Thank you

And to my dear grandmother, Nelida Maldez, who passed away this past year,

Lela Nelly, un beso grandote de tu gaucho hilacha.

Chapter 1

INTRODUCTION

This investigation is first and foremost about context. What is context? Tessmer and Richey (1997) define context as "…a multilevel body of factors in which learning and performance are embedded" (p. 87). These situational factors can include the time, place, emotional states, sights, sounds, smells, and historical conditions in which learning and performance occur. For example, the reader is currently perusing the words in this document, and although the reader's attention is focused on the words and their meaning, the reader also perceives a variety of sensory stimuli (sights, sounds, and smells). Moreover, the reader's processing of these words is taking place in parallel with internal emotional states that have not necessarily been caused by reading this document. In other words, reading these words is taking place within a broader context, consisting of a myriad of internal and external situational factors.

Secondly, this study is about the presence of color. Color is an omnipresent screen design variable that, among other things, can help to link related content and to communicate the over-arching content structure to the learner (Milheim & Lavix, 1992). Winn and Everett (1979) support the use of color as a cuing device and suggest that color improves a learner's ability to recall both relevant and irrelevant material. Pett and Wilson (1996) suggest that colors can be used to code content and to link it logically to related elements.

Thirdly, this study is about improving memory recall. *Recall* is a form of retrieval that pertains to an individual's ability to reconstruct a stored event or episode. This is in contrast to *recognition*, which is a form of retrieval pertaining to an individual's ability to identify something as having previously been experienced. Why is recall an important learning outcome? In some professional settings, such as in medicine and law, the practitioner's credibility is highly dependent upon his/her ability to recall information. For example, how credible would a trial lawyer be if, in the midst of a legal argument in court, the attorney were unable to recall the name of a prior precedent-setting case? How much confidence would one have in a surgeon who could not remember the name of the various instruments and medications needed during a procedure? Imagine a surgeon saying to a nurse, "Give me 3 CCs of that clear liquid that reduces clotting and the sharp metal knife-like instrument there on the tray." Being able to recognize an instrument and even being able use it correctly are not enough. The surgeon must be able to recall the correct name in order to communicate accurately, precisely, and efficiently. The ability to recall information is no less critical in other professions and in other settings, e.g., an eye-witness providing testimony, a pilot recalling crash landing procedures, or a server remembering a customer's order. Clearly, recall is an important learning and performance ingredient. It seems self-evident that instructional designers and educators should be concerned with what variables they can manipulate in order to improve an individual's ability to recall information.

The current investigation examines the effects of a single context variable, the presence, or absence, of border colors on an individual's ability to recall factual and conceptual information.

Types of Context

For the purposes of the current study, context refers to any stimulus that can be perceived (seen, heard, felt, tasted, or smelled) that is *not* intended to be the focus of the learner's interest. In educational settings, this can include almost any factor except for the actual to-be-remembered information presented. Context variables can be classified as global or local and as internal or external (Mori & Graf, 1996).

Global context. Global, or environmental, context variables are relatively easy to identify and include such variables as odors, background noise, the feel of a seat, the temperature of a room, and the decorations on the wall. Continuing the previous example, the reader may be scanning this document in his or her home. Different rooms in the house can have different lighting, different color paint or wallpaper, and different furniture… all contextual factors that combine to give each room a different *look and feel*. All of these broad environmental factors are categorized as global context variables. The effects of global context cues on learning have been well documented (see Baddeley, 1999).

Local context. Local context variables, however, are more subtle than global ones insofar as the former are proximal to the object of learning and can often commingle with it. For example, the color of the paper on which the words are written could be a local context variable. If this document were printed on blue paper, or if each page were printed on paper of a different hue, the reader could not help but see the color as he or she was reading the text. On the other hand, the color of the wallpaper in the room would not be proximal to the words and would therefore be considered global. It is this proximity that distinguishes global from local context.

Local, internal context. Local, internal context variables are those that are not only near the object of learning, but rather part of it. They are intentionally and systematically commingled with the words the reader is trying to understand. The background color of the paper, the size and type of font used, the presence of color highlighting of certain key terms, as well as the bolding and italicizing of others are all local contextual factors that could be embedded in the words on this page. An instructional designer might, for example, create an illustration depicting the various components of a car's engine. Each component of that engine is depicted using a different color hue. The text accompanying the illustration might be shown in different colors as well, in such a way as to correspond with the different engine components. Green text could correspond with an illustration of the carburetor drawn in green, red text with the electrical system, and so on with other components. The manipulation of color in the illustration and the text is clearly intended for the learner to focus on the color associations as well as the to-be-remembered content. The intent is that the color associations will help distinguish the different components and promote stronger clustering of content in memory.

Local, external context. Local, external context variables exist somewhere in between global factors (such as room temperature) and local, internal factors (such as text highlighting) and have not been researched as extensively in educational settings. Local, external context cues are not intentionally commingled with the to-be-remembered (TBR) content and yet they are adjacent to the TBR content. In a study involving color as a local, external context variable, Clariana (2002) manipulated the color of the screen's navigation bar. The navigation bar was adjacent to the content area of the screen, but the content area contained black text on a white background. The colors used were not commingled with the TBR content, and yet the results suggest that color information was nonetheless encoded. In addition, when those colors were shown again at posttest, they helped learners remember more of the TBR content. This suggests that even when color is not commingled with the intended focus of the instruction, it can be used to facilitate learning. Given the extensive use of color in computer-based instruction, screen design, and user interface design, such findings could have widespread application.

Statement of the Problem

The study of contextual factors that influence learning and performance is promoted by a number of leading instructional design researchers (Dick & Carey, 1990; Jonassen, 1993; Richey, 1992; Smith & Ragan, 1993; Tessmer & Harris, 1992). In the present study, a seemingly innocuous external and local context variable, the border color of a screen in a course presented in the form of computer-based instruction, is examined for its effects on recall.

Although cognitive research involving paired-associate learning suggests that context cues, including color, can have a powerful effect on memory (Baddeley, 1999; Elio & Reutener, 1978; Godden & Baddeley, 1975, 1980; Murnane & Phelps, 1993, 1994, 1995; Murnane, Phelps, & Malmberg, 1999; Pellegrino & Salzberg, 1975; Tessmer & Richey, 1997), there is little empirical guidance for how such cues can be used to improve recall in realistic instructional settings. There is even less direction for how color might be used to enhance recall in computer-based instruction, despite its ubiquity as a screen design element.

Paired-associate research

Paired-associate research, while it presents a number of significant insights into how memory processes function, is nonetheless incapable of providing practical guidelines for how context cues can be used to improve retrieval in real-life situations. In discussing the commonly used methods of testing memory, Morris (1978) states:

> The permanent problem facing the psychologist studying memory is the need to test situations sufficiently similar to everyday life for the results to be generalizable beyond the laboratory while at the same time retaining sufficient control so that accurate scoring can reveal what differences his manipulations of the variables have made. (p. 27)

Morris goes on to describe serial learning and paired-associate learning as "…highly artificial…" suggesting that in spite of that, "…they may to some degree simulate fundamental features of memory, such as the specific cues to elicit recall" (p.28). Nevertheless, if memory research is ultimately going to be useful for educational practitioners, it is important that its findings be applied to real-life educational settings. This study attempts to do just that by applying the findings of paired-associate list learning research to substantive educational materials presented in the form of computer-based instruction.

Print-based color research

While some educational studies have involved the use of color in print-based materials (see Lamberski & Dwyer, 1983; Moore & Dwyer, 1997), relatively few studies have specifically addressed it in the context of computer-based instruction (Clariana, 2002; Williams, 2000). Several educational researchers caution against the wholesale application of research findings from print-based to screen-based media (Kolers, Duchnicky, & Ferguson, 1981; Schwier & Misanchuk, 1993), suggesting that print-based findings must be tested in screen-based settings.

Local, external context cues

Most color studies naturally involve local, *internal* uses of color, such as color highlighted text (Worley & Moore, 2001). In these studies, color is intentionally intermingled with TBR information. With studies involving local, *external* context cues, such as headings, motifs, navigation bars, and border colors, the contextual information is usually not intermingled with the to-be-remembered (TBR) information. The category headings context effects found by Epstein and his colleagues (Epstein & Dupree, 1977; Epstein, Dupree, & Gronikowski, 1979) and navigation bar context effects found by Clariana (2002) suggest that local, *external* context effects should be investigated further. Lamberski and Dwyer (1981) write:

> Apparently, what is needed is additional systematic research designed to effect a more comprehensive assessment of the instructional potential inherent in color and the ways it might optimally be employed to improve student achievement of designated educational outcomes. (p. 130)

Summary. Although a large number of memory research and instructional design research studies have investigated the effects of context cues on learning, paired-associate list learning research findings may not be generalizable to more realistic instructional settings. Also, print-based findings may not be generalizable to computer-based instruction. In addition, neither body of research has fully addressed questions related to the effectiveness of local, external context cues, such as border color. Therefore, further investigation is warranted.

Justification for the Study

A considerable amount of educational research has been dedicated to studying color, with mixed results. Chute (1979) suggests that color research in educational settings has generally yielded no significant differences in terms of posttest achievement. He writes:

> The effectiveness of color cuing in media presentations has been researched for more than 30 years. However, the findings of many of these research efforts have been inconclusive; consequently, they have provided few guidelines to educational designers concerning the use of color cuing in media presentations. In general, the research indicates that while learners prefer color versions to black-and-white versions of media presentations, they do not consistently learn significantly better from the color versions. (p. 251)

Why after decades of finding no significant differences would researchers now expect to find something different? Chute suggests that one reason color research has failed to find significant differences in learning achievement is that past research has not taken into account the influence of individual differences. For example, researchers have found that high-ability learners are able to process more sensory information than low ability learners and that low ability learners take longer and require more highly structured information (Cronbach & Snow, 1977). This suggests that color could influence high-ability and low-ability learners differently.

When a designer chooses to commingle color with to-be-remembered (TBR) text and/or illustrations, the designer relies on the learner to 1) perceive the presence of color, 2) distinguish it as more than just visual clutter, 3) understand the color-coding scheme that the designer is using to communicate the structural relationship between color and content, and 4) make use of the color-coding information at retrieval. A wide variety of factors can influence each of these cognitive tasks and thereby interfere with research results. For one thing, the associative benefits of color-coding instructional materials may be outweighed by the incremental memory load (Miller, 1956; Sweller, 1988) imposed on the learner from processing the color information actively in parallel with the TBR content. Also, individual differences in cognitive ability, cognitive style, and simple preference could conceivably influence how/whether color is encoded and used, thereby inhibiting or enhancing the effects of color. Dwyer and Moore (1994), for example, found that color highlighting of text helped field *dependent* learners but not field *independent* learners.

In addition to individual differences, consider the role of item differences, i.e., differences in the nature of the mental tasks that learners are asked to complete. Cognitive research has consistently found that context cues do *not* influence *recognition* retrieval tasks, even if those same cues sometimes influence *recall* retrieval tasks

(Baddeley, 1999). Therefore, to the extent that educational researchers use multiple-choice, matching, and other recognition test formats, they may not consistently find significant color effects.

With individual and item differences in mind, this and similar studies (Chute, 1980; Clariana, 2002, 2003; Dwyer & Moore, 1994; Dwyer & Moore, 2001; Lamberski & Dwyer, 1981, 1983; Ling & Blades, 1996) are infusing new life into the investigation of color effects on learning.

Purpose of the Study

The main purpose of this investigation was to examine the relative effects of contextual border colors presented in a computer-based lesson (at encoding) and/or in a fill-in-the-blank posttest (at retrieval). The posttest was designed to assess recall of factual and conceptual knowledge. The display of border color was manipulated as a local, external context cue. Within this framework, the current investigation was designed to:

1. Determine whether or not presenting border colors in the lesson and/or the posttest improved recall of factual and conceptual knowledge.
2. Determine whether or not color context effects were influenced by item type, item difficulty, and/or gender.

Research Questions

Following are the research questions examined in this investigation. Each question pertains to recall task performance and involves undergraduate students completing a computer-based tutorial and posttest.

1. Does the presence of border colors in the lesson or posttest enhance recall of factual and conceptual knowledge?
2. Does *matching* the border color displayed in the posttest with that displayed in the lesson optimize recall of factual and conceptual knowledge?
3. Is there an interaction between the type of factual recall task assessed (involving verbatim and paraphrased/transposed test items) in the posttest and the effects of border color on recall of factual and conceptual knowledge?
4. Is there an interaction between posttest item difficulty (easy and difficult) and the effects of border color on recall of factual and conceptual knowledge?
5. Does recall of factual knowledge vary from that of higher-level conceptual knowledge in terms of the effects of border color?
6. Is there an interaction between gender and the effects of border color on recall of factual and conceptual knowledge?
7. Does *mismatching* the border color displayed in the posttest with that displayed in the lesson undermine recall of factual and conceptual knowledge?

Research Hypotheses

Alternate hypotheses	Null hypotheses
H_1 Students viewing lesson or posttest materials with border colors will recall significantly more factual and conceptual information than students viewing materials with no border colors.	H_0 The presence of border colors in the lesson or posttest materials will have no significant effect on recall of factual and conceptual knowledge.
H_2 The effects of color on recall of factual and conceptual knowledge will be strongest when there are border colors in the lesson and matching border colors in the posttest.	H_0 The effects of color on recall of factual and conceptual knowledge will be the same regardless of whether or not the color is present in the lesson, the posttest, or both.
H_3 There will be a significant interaction between the type of recall task performed (involving verbatim and paraphrased/transposed test items) and the effects of border color on recall of factual and conceptual knowledge.	H_0 There will be no significant interaction between the type of recall task performed and the effects of border color on recall of factual and conceptual knowledge.
H_4 There will be a significant interaction between item difficulty (easy and difficult) and the effects of border color on recall of factual and conceptual knowledge.	H_0 There will be no significant interaction between item difficulty (easy and difficult) and the effects of border color on recall of factual and conceptual knowledge.
H_5 There will be a significant interaction between the knowledge level assessed (factual and conceptual knowledge) and the effects of border color on recall of factual and conceptual knowledge.	H_0 There will be no significant interaction between the knowledge level assessed (factual and conceptual knowledge) and the effects of border color on recall of factual and conceptual knowledge.
H_6 There will be a significant interaction between gender and the effects of border color on recall of factual and conceptual knowledge.	H_0 There will be no significant interaction between gender and the effects of border color on recall of factual and conceptual knowledge.

H_7	H_0
Mismatching the lesson's border colors with the border colors used in the posttest will inhibit recall of factual and conceptual knowledge and therefore yield negative effect sizes relative to the color-matching group.	The color-mismatching group will recall no less factual and conceptual knowledge than the color-matching group.

Generalizability

The findings of this study may be generalized in the following manner:

1. Since the students completed a computer-based instruction module involving factual learning of abstract concepts, the results may be generalized to similar forms of computer-based instruction, delivered via CD-ROM or over web connections, in which no practice is provided.
2. Since the students involved in this study were Introduction to Psychology undergraduate students at a major northeastern university, the results may be generalized to similar undergraduate students enrolled in general education courses for school credit at similar U.S. institutions.
3. The results may be generalized to situations in which the learning objectives call for the recall of factual and conceptual knowledge of abstract concepts.
4. The results of the current study were based on the use of 8 randomly selected color hues. Since different color hues can have different effects on an individual's emotional state, such as blue eliciting calm and red eliciting excitement (Bellizzi & Hite, 1992), these results should be generalized to similar combinations of hues.

Limitations of the Study

Generalization of these research findings should be restrained in the following manner:

1. The results should not be generalized to populations outside of undergraduate students. Younger children may possess different cognitive abilities/constraints that enable/hinder their ability to make use of contextual variables. Older adults may differ as well.
2. Since the participants experienced no audio, video, or animations, and to the extent that these additional media may compete with color context effects, the results should not be generalized to more robust multimedia presentations.
3. Similarly, these results should not be generalized to forms of computer-based instruction that involve practice activities, such as tutorials. The instructional materials used in the current study involve participants reading instructional text and provide no opportunity for practice.
4. The results should not be generalized to instructional outcomes involving psychomotor, affective, or high-level cognitive skills. These results apply only to recall of factual knowledge and conceptual knowledge.

Definitions

Color-coding. This refers to the strategy of using different colors to designate different sub-units of content in order to cue the overarching structure of the content. In this study, color-coding was used to differentiate eight instructional lessons. The colors were used in the border around the screen and in a decorative swirl that runs down the right side of the screen.

Context effect. Context effect refers to a contextual cue's ability to influence the retrieval of previously learned (encoded) information. Typical context effect studies compare experimental conditions in which context cues are presented during the learning task (encoding), the retrieval task, or both against a control group in which no cues are presented. The difference between the cued condition and the uncued control group condition represents the context effect.

Cuing Conditions. A cuing condition refers to the manipulation of the context cue in the treatments. In paired-associate learning, a context variable is either shown (cued = C) or not shown (uncued = U) during encoding and retrieval. In a typical study, there are four cuing conditions: (1) U-U, the control group in which the context variable is not active at all; (2) U-C, the context variable is not shown during encoding but is shown at retrieval; (3) C-U, the context variable is shown during encoding but not during retrieval; and (4) C-C, the context variable is active both during encoding and retrieval. In the present study, border colors are the context cues being manipulated. Instead of using *U* (uncued) and *C* (cued) to describe the cuing conditions for the present investigation, *NoC* (no color) and *C* (color) were used. Also, instead of using the terms *encoding* and *retrieval*, the terms *lesson* and *posttest* were substituted. These conventions were used for the sake of clarity, as they represent terminology more commonly seen in instructional design research.

Encoding tasks. Baine (1986) describes encoding as "…perceptual processing…whereby sensory stimuli are converted, analyzed, and stored in memory" (p.

10). Encoding can be verbal or non-verbal (e.g., structural and visual). In this study, the explicit encoding task for the participants was reading a passage containing to-be-remembered factual and conceptual information, a verbal encoding task.

Episodic memory. According to Tulving and Thomson (1973), episodic memory "is concerned with storage and retrieval of temporally dated, spatially located, and externally experienced events or episodes, and temporal-spatial relations among such events" (p. 354). In other words, episodic memories are embedded in the context of our experiences and are stored much like a collection of movies. Contextual information is, in a figurative sense, the background that provides texture for those episodes.

Implicit memory. In a review of implicit memory literature, Schracter (1987) distinguishes implicit memory as being "revealed when previous experiences facilitate performance on a task that does not require conscious or intentional recollection of those experiences," while explicit memory "is revealed when performance on a task requires conscious recollection of previous experiences" (p. 501).

Recall. Retrieval, recall, and recognition are often used synonymously to describe memory, yet each is distinct. Recall, a type of memory retrieval, refers to an individual's ability to reconstruct a stored event or episode. Recall is typically assessed through cued recall and free recall tasks. With cued recall, the individual is prompted with relevant information about the original event or episode. For example, in paired-associate list learning, the individual is given a list of paired terms and then tested by being given the first term and being asked to recall the second term. With free recall, the individual is asked to remember the terms in whatever order he/she can remember them, i.e., without the first term to cue memory. There are several variations of these approaches. In this study, fill-in-the-blank questions were used to test cued recall.

Recognition. Recognition, also a form of memory retrieval, refers to an individual's ability to identify something as having previously been perceived or learned.

Retrieval. Retrieval is a broad term used to describe the act of remembering information that has previously been encoded in memory. Retrieval can take the form of recall or recognition.

Semantic memory. According to Tulving and Thomson (1973), semantic memory is "…the system concerned with storage and utilization of knowledge about words and concepts, their properties, and interrelationships" (p. 354). In other words, semantic memory is concerned with the conceptual understanding of what we remember. This is comparable to what Brainerd and Reyna call gist traces (1990), which capture the meaning of the stimuli we experience.

TBR Content. TBR, or to-be-remembered, content refers to the information presented in the instructional lesson that has been identified as the information that needs to be learned. It is intended to be the object of the learner's immediate attention. In this study, the TBR content refers to the definitions and concepts that study participants were asked to read and learn.

Trace (Memory trace). According to Tulving and Thompson (1973), a memory trace refers to "…a multidimensional collection of elements, features, or attitudes…" (p. 354) that is perceived by the sensory system and encoded as memory. A trace is the basic unit of memory.

Chapter 2

REVIEW OF THE LITERATURE

The main purpose of this investigation was to examine the relative context effects of border colors presented in a computer-based lesson (at encoding) and/or in a fill-in-the-blank posttest (at retrieval). The posttest was designed to assess recall of factual and conceptual knowledge. The display of border color was manipulated as a local, external context cue. Within this framework, the current investigation was designed to address the following questions:

1. Does the presence of border colors in the lesson or posttest enhance recall task performance?
2. Does *matching* the border color displayed in the posttest with that displayed in the lesson optimize posttest achievement?
3. Is there an interaction between the type of factual recall task assessed (involving verbatim and paraphrased/transposed test items) in the posttest and the effects of border color?
4. Is there an interaction between posttest item difficulty (easy and difficult) and the effects of border color?
5. Does recall of factual knowledge vary from that of higher-level conceptual knowledge in terms of the effects of color?
6. Is there an interaction between gender and the effects of color?
7. Does *mismatching* the border color displayed in the posttest with that displayed in the lesson undermine posttest achievement?

Given these research questions, this review of literature addresses research in the following areas: color, context dependency, the differential effects of item types and item difficulty, and gender-related cognitive differences.

Color research suggests that color may be encoded automatically with to-be-remembered content but that it may not be a very potent variable in learning. Context dependency research, specifically the *encoding specificity principle*, suggests that cueing retrieval with a context variable introduced during encoding may enhance recall. With regard to item difficulty, the *outshining hypothesis* (Smith, 1988) predicts that more difficult test items will, in the absence of better cues, be more context-dependent than easy items (p. 19). Gender research suggests that there will be differences favoring females in their ability to recall verbal information (Weaver & Raptis, 2001). All of these areas will be examined more closely in this review of literature.

Context Dependency

Encoding specificity principle

Tulving posits that context cues can only impact retrieval when they are available both at encoding and retrieval (Tulving, 1972). In a series of experiments, Tulving and his associates (Thomson & Tulving, 1970; Tulving, 1972; Tulving & Osler, 1968; Tulving & Thomson, 1973) found that associated words presented both at encoding and retrieval cued recall of the TBR words, whereas they did not cue recall when presented at encoding or retrieval alone. Tulving and Thomson (1973) write: "…only that can be retrieved that has been stored, and how it can be retrieved depends on how it was stored…" (p. 359). For example, Dulsky (1935) found that when one background color was presented at encoding and then a different background color was presented at retrieval, recall of TBR words was suppressed. Using verbal cues rather than visual ones, Tulving and Thomson (1973) found that the presence or absence of context cues at retrieval impacted recall of TBR words. In other words, a cue must be present at the time the material is learned in order for it to be an effective memory retrieval aid.

Although Tulving's findings have been reproduced using a variety of context cues, it should be emphasized that in his studies, Tulving used paired-associate words as internal (intralist) and external (extralist) verbal context cues, not color or other nonverbal cues. In a study involving the presentation of one or two weakly associated words (the context cues) during encoding and retrieval, Tulving and Osler (1968) showed 278 Canadian high school students a list containing 24 TBR words. Table 1 depicts the general 2 x 2 crossed design of Tulving's studies. In the cued encoding conditions (C-U and C-C), the TBR words were shown alongside two weakly associated words. Here are some examples with the TBR words capitalized and the paired-associate words in lower-case letters: *fat, leg – MUTTON; village, dirty – CITY; and dark, girl – SHORT*. In the uncued encoding conditions (U-C and U-U), only the TBR words were shown. At retrieval, the learner was either cued with the paired-associate words adjacent to each blank space or was not cued at all (only blanks spaces appeared on the test).

Tulving and Osler's (1968) results indicated that the number of cues (one word or two) made little difference. More importantly, only when the associated words were present at *both* encoding and retrieval (i.e., in the matching C-C condition) were significant recall context effects found. Context effects refer to the difference between each experimental condition (C-C, C-U, and U-C) and the control condition (U-U). In this example, the two weakly associated words were the context cues (Tulving & Osler, 1968). In other experiments, Tulving and his associates used both weakly and strongly associated words (such as ground, hot – COLD) (Thomson & Tulving, 1970), extralist, congruous, and incongruous words (such as tennis, player, whistle – BALL) (Tulving, 1974), as well as related and unrelated rhyming words (such as desk, pair – CHAIR) (Tulving & Watkins, 1975). In all of these studies, Tulving and his colleagues found that the context cue was only effective when it was present both at encoding and retrieval. In other words, the C-C condition is the only one in which there were context effects.

Table 1

Prototypical study design for encoding specificity studies

Input (encoding) condition	Output (retrieval) condition	
	Uncued	Cued
Uncued	U-U	U-C
Cued	C-U	C-C

Each of the studies listed above supports the encoding specificity principle, which predicts that context cues will improve retrieval *only* if they are made available both during the lesson (encoding) and during the posttest (retrieval). Otherwise, context cues should have no effect when presented during the lesson only or during the posttest only. A strict interpretation of encoding specificity suggests that both U-U and C-C conditions would promote retrieval (the difference between the two representing the context effect of the cue), while the U-C and C-U conditions would suppress retrieval, since the trace link connecting the context with the TBR content is missing in both conditions. Such a strict interpretation, however, may not be appropriate when other types of context cues are employed.

Generation-recognition model

When they presented the encoding specificity principle, Tulving and Thomson (1973) challenged other contemporary theories of memory, in particular, the generation-recognition model. Since then, supporters of the latter have countered with theoretical challenges (see Martin, 1975; Santa & Lamwers, 1974, 1975) and have chipped away at the encoding specificity principle with empirical findings that limit its generalizability (Epstein & Dupree, 1977; Epstein et al., 1979).

Recognition models of memory date back to the industrial revolution (James, 1890) and generally hold that in trying to retrieve TBR information, one always utilizes an approach that substantively resembles the process of recognition. Bahrick (1970) suggested that recall involves a two-step process of generating alternatives and recognizing the correct information. Consider the process one might follow if attending a high school reunion and asked to remember the name of an old friend. If the name does not immediately spring to mind, one might initially generate a mental list of possible names; then, through an automatic process of elimination and recognition arrive at the old friend's name (Kintsch, 1970). If the old friend were standing across the room, one might utilize the available contextual cues (e.g., physical appearance, voice, manners of speech, etc.) to inform one's generation process and thereby restrict the number of options to consider (Bahrick, 1970). Otherwise, proponents of the generation-recognition model do not distinguish between the mental processes of recognition and recall (Epstein

& Dupree, 1977). With the generation-recognition model, context cues are relegated to being used only to restrict the generated lists, facilitating faster recognition and reducing the chances of recognizing the wrong option, and are not considered essential for retrieval (Tulving & Thomson, 1973). Conversely, in the eyes of Tulving and supporters of the encoding specificity principle, recall and recognition are fundamentally different mental tasks, and make different uses of the contextual information available.

Challenges to the encoding specificity principle

Critics of the encoding specificity principle have challenged Tulving and his associates through a number of different studies. In two experiments involving the use of category headings as encoding and retrieval cues, Epstein and Dupree (1977) found that although C-C participants, who were cued with headings both at encoding *and* at retrieval, scored highest on a cued recall task, the U-C and C-U groups, in which headings were provided at retrieval only and at encoding only (respectively), performed significantly better than the (U-U) control group also. The encoding specificity principle would have predicted that *only* the C-C participants would have performed well. In this study, the U-C group performed just as well as the C-C group. They explain this by writing: "… the lack of cuing at recall depressed performance greatly if the items had also not been categorized at input and depressed performance only slightly if the items had been initially categorized" (p.105). Such findings favor the generation-recognition model, which suggests that regardless of where the context cue is placed, it plays a secondary role. Therefore, this model predicts that U-C, C-U, and C-C conditions will have equivalent levels of recall.

Epstein and Dupree (1977) also found that when context cues were given at encoding but not at retrieval (C-U), participants performed significantly better in *free* recall tasks than those who were not cued at all (U-C). They wrote: "Thus, it is seen that in the absence of explicit recall cues, free recall is greatly enhanced by imposed organization during acquisition" (p. 104). As Epstein and Dupree (1977) point out, encoding specificity does not account for such organizational effects.

While their earlier study appeared to challenge the encoding specificity principle, Epstein et al.'s (1979) study failed to support their hypothesis that the context effects of headings could be predicted by the generation-recognition model. In this set of experiments, the researchers replicated their earlier study but also manipulated the number of headings categories (2, 3, 5 and 6 categories). This time, even though cuing improved recall when applied at input, output, and both input and output, they did find significant differences between the C-C and the C-U conditions (contrary to their earlier study). From these results, they concluded that their study of category headings as contextual cues could neither support their earlier generation-recognition hypothesis (since C-C did produce significantly greater recall scores than all other cuing strategies), nor could it support a strict interpretation of Tulving's encoding specificity principle (since the other cuing strategies U-C and C-U also produced significantly higher recall scores than the U-U control group).

Epstein et al. (1979) also found that the context effects of headings declined as more categories were added. It is possible that – as Epstein, Dupree, and Gronikowski (1979) suggest – the exploration of "situationally specific mini-theories" (p. 180) for contextually cued retrieval are necessary to explain the effects of various types of context cues as they interact with other instructional variables, such as encoding strategies. For example, the effects found when headings were introduced at encoding but withheld at retrieval (C-U) (Epstein & Dupree, 1977) can be partially explained by the advanced organizing effects of logical heading structures (Ausubel, 1968). Similarly, the small context effects that Epstein and his colleagues found for headings presented during retrieval but not during encoding (U-C) (Epstein & Dupree, 1977; Epstein et al., 1979) could be explained in terms of attention-arousal motivation effects (Keller, 1999; Keller & Kopp, 1987).

In an effort to reconcile encoding specificity with generation-recognition, Pellegrino and Salzberg (1975) suggest a less strict interpretation of encoding specificity. They posit: "…the subject is *not* bound to incorrect responding if the cue or context is changed" (p. 270). This interpretation suggests that while the C-C condition may enhance retrieval the most (as evidenced by recall score differences between the U-U and C-C conditions), it does not necessarily mean that cue incompatibility will always suppress scores relative to the U-U condition. According to Epstein et al. (1979), Tulving conceded to this more relaxed interpretation of the encoding specificity principle in a personal communication. Thus, with this revision, encoding specificity predicts that while context cues may impact recall when presented at either encoding (C-U) *or* retrieval (U-C), they should have the greatest impact on recall when presented in both encoding *and* retrieval (C-C).

Thus, for the purposes of the present experiment, the encoding specificity principle predicts that in the homogeneous C-C condition, in which color is present in the lesson *and* the posttest, recall will be optimized. The heterogeneous conditions (U-C and

C-U) may enhance recall also, but not as much. Therefore, the C-C group should recall significantly more TBR information than the other conditions. The generation-recognition model, on the other hand, predicts that all cued conditions (U-C, C-U, and C-C) will enhance recall equally well.

Color in Context

Color is a common element in the visual array. When color is presented in instructional materials, it is non-verbally encoded into memory as visual stimuli. Hasher and Zacks' (1979) automaticity hypothesis posits that encoding lies on a continuum of effortful to automatic, and that context variables, such as color, tend to lie on the automatic end of this continuum. In other words, regardless of whether or not the learner is trying to remember color, he or she will encode color automatically, without any additional effort. A study by Patel, Blades, and Andrade (1999) supports this hypothesis. Eighty participants of various ages (4-, 7-, and 10-year olds as well as adults) were asked to study eight different objects of different colors (i.e., green comb, orange pen, brown cup, white book, red sock, black bowl, blue ball, and yellow toothbrush) for an equal amount of time. Note that in this study, color was manipulated as a local, internal context variable. The color was commingled with the TBR content, i.e., the colored objects. Thirty minutes after the study task, the participants were given a surprise memory test. The test assessed both recall and recognition of color, not recall or recognition of the items (the TBR content). Patel et al. (1999) reported that participants on average remembered correctly the colors for six of the objects (6 of 8, or 75% correct). The results indicated that participants of all ages and gender performed significantly better than would be predicted by random guessing (1 out of 8, or 12.5% correct). This was the case for both primary colors and secondary colors. These results suggest that even when learners are not asked to remember color, color is nonetheless encoded automatically and stored for later retrieval. Other studies support the notion that color is learned automatically (Backman, Nilsson, & Nouri, 1993; Hatwell, 1995; Ling & Blades, 1996). Taken as a whole, these cognitive studies suggest that color is indeed stored in memory even when the learner is not making any effort to remember color.

According to Thompson (1994) color is used by instructional designers to give visuals a sense of realism, i.e., to make them "…more concrete representations…" (p. 171); to evoke emotional responses from the learner (e.g., red for excitement); and to "…differentiate different portions of a visual…" (p. 171) in the form of color-coding. Some studies have shown that using color as a coding device can be an effective strategy for promoting both list learning and concept learning (see Dwyer, 1967; Dwyer & Lamberski, 1982-83 for reviews).

In a study involving 132 undergraduate students, Dwyer and Moore (1994) found that students who read instructional materials containing color-coded line drawings performed significantly better on verbal test items than students who read materials with black-and-white coded drawings. In this case, color was used as a local, internal context variable to "…highlight (the) information and processes being presented…" (p. 218). On the other hand, Chute (1979) and others (see Otto & Askov, 1968) have suggested that color is susceptible to being over-powered by other cues, particularly ones pertaining to form (such as headings perhaps).

In a study involving color-coded print-based materials and a series of four posttests (3 recognition tests and 1 recall test) measuring different learning outcomes, Lamberski and Dwyer (1983) found that an "…integrated color code provides the greatest potential at retrieval if available and utilized during concept acquisition or encoding…" (p. 18), not necessarily when it is made available at retrieval (Note: recall was assessed using a drawing test). In other words, participants performed better when color was given in the lesson but not when it was present in the posttests. Their findings and interpretations run counter to the long-debated encoding specificity principle (Tulving & Thomson, 1973), but are consistent with the generation-recognition model (Bahrick, 1970; Epstein & Dupree, 1977; Martin, 1975; Santa & Lamwers, 1974).

Canelos, Taylor, and Dwyer (1985) applied encoding specificity to a study of visualized instruction in a real-life classroom setting. In this experiment, the encoding strategy (*visually*-mediated versus *verbally*-mediated instruction) and retrieval strategy (*free* recall, *visually cued* recall, and *verbally cued* recall) were manipulated to form six treatment groups. With verbalized mediated instruction, the results were generally consistent with encoding specificity insofar as verbally cued participants (M = 15.18) significantly outscored visually cued ones (M = 6.93). The homogeneous C-C condition was by far more effective than the mismatching condition. However, with visually mediated instruction, the visually cued participants (M = 15.30) did not score significantly higher than the verbally cued ones (M = 13.68). Although the homogeneous C-C condition produced higher achievement than the control U-U condition, the heterogeneous U-C and C-U conditions were also high for visually-mediated instruction.

While the former finding is consistent with encoding specificity, the latter brings the strict interpretation into doubt.

Such findings suggest that using words to cue other words, as in paired-associate research, may involve a fundamentally different mental process than using images, highlighting, and other nonverbal context variables to cue words. Canelos et al.'s (1985) study also suggests that verbal cues (text) may be more potent than visual cues (such as graphics and color) regardless of whether the instruction is verbally or visually oriented.

Context effect differences by item difficulty

Different mental tasks place different cognitive demands on students (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956; Nitko, 1996) and it is possible that, in doing so, different tasks cause students to depend more or less on contextual cues for help. The literature suggests that both item difficulty and the desired educational outcome level of tasks may be related to context dependency.

In a recent study involving color-coded vertical bars mimicking the navigation bars of web pages, Clariana (2002) found that color interacted with fill-in practice exercises to enhance recall for difficult constructed response test items but failed to produce context effects for recognition tasks or easy recall tasks (see also Clariana, in press). Similarly, Worley and Moore (2001) found that highlight color, an internal context cue, did not improve recognition task performance on a criterion test. These findings are consistent with paired-associate learning research indicating that *global* context cues have a significant effect on recall tasks while having no substantive impact on recognition tasks (See Godden & Baddeley, 1980; Jacoby, 1974, 1983; Kintsch, 1970;

McCormack, 1972; Smith, Glenberg, & Bjork, 1978; Weiss & Margolius, 1954). In general, it appears that different retrieval tasks may differ in terms of their context dependency.

Outshining hypothesis

The differential effects of context variables upon recognition and recall may be explained by the outshining hypothesis. The outshining hypothesis, proposed by Smith and his colleagues (Smith, 1986; Smith, 1988, 1994; Smith et al., 1978), generally states that the trace activation strength during recognition tasks is much greater than it is during recall tasks; therefore, the TBR content stored in memory outshines the stored context information. On the other hand, the retrieval activation strength during recall tasks is generally lower; therefore, context information has greater opportunity to assist in the retrieval of the TBR content. Another way of thinking of this is that with recognition tasks, the match is so strong (the retrieval task so easy) that the learner does not need to search for context information that might clue them into the answer. This hypothesis is relevant to the present study because it suggests that both the nature of the retrieval task (test or item format) and the difficulty of the task are inter-related and impact context dependency. Therefore, their influence on context effects should be considered separately and yet at the same time.

Insofar as both the type of retrieval task and the difficulty of that task are related to the activation strength of the TBR content, the role of context in retrieval tasks may vary. For example, the outshining hypothesis predicts that a more difficult retrieval task involving recall involves TBR information with low activation strength and therefore the learner will be more dependent on context. An easy recall task would involve TBR content with higher activation strength and therefore the context effects would be lower. On the other extreme, easy recognition tasks have such activation strength that context cues may not be required at all. These predictions are consistent with Clariana's (2002) findings that color helped students only with difficult recall tasks.

In summary, the outshining hypothesis suggests that the more difficult the TBR content is to retrieve from memory, the more reliant the learner is upon contextual cues. For the purposes of the current study involving fill-in-the-blank recall tasks, the outshining hypothesis predicts that learners will benefit from color context cues more when they are faced with difficult posttest items than when they must complete relatively easy posttest items.

Taxonomies of educational objectives

In their *Taxonomy of Educational Objectives* (1956), Bloom and his colleagues defined six classifications: knowledge, comprehension, application, analysis, synthesis, and evaluation. Subsequently, alternative taxonomies have emerged. Gagnè's (1968) taxonomy contains eight categories: *information*, *concrete concepts*, *defined concepts*, *principles*, *problem solving*, *cognitive strategies*, *attitudes*, and *motor skills*. Merrill's (1999) taxonomy emphasizes *facts*, *procedures*, *concepts*, *rules*, and *principles*. Most taxonomies have in common the characteristic of being one-dimensional. That is, in categorizing objectives, they each combine the nature of the content with the nature of the task. For example, according to Gagnè's taxonomy, *information*, *concepts*, and *principles* are types of knowledge, while *problem solving* is a type of task, yet all of these are listed in a one-dimensional taxonomy. Similarly, in Bloom's taxonomy, the knowledge category is composed of *knowledge of specifics*, *knowledge of ways and means*, and *knowledge of universals and abstractions*. Knowledge is classified by the nature of the TBR content. On the other hand, *application* is characterized by the cognitive task involved (applying information to a novel context), not the nature of the knowledge itself. This mixing of content type and cognitive task, noun and verb, is somewhat limiting (Krathwohl, 2002).

Recently, Anderson, Krathwohl, and several other distinguished leaders in the field (2001) published a revision of Bloom's original taxonomy (see Table 2). In it, the authors propose a two-dimensional model for categorizing educational objectives. In one dimension, they classify the different types of knowledge (nouns) from simple to complex, from concrete to abstract: factual, conceptual, procedural, and metacognitive knowledge. In the present study, both factual and conceptual knowledge were assessed. In the revised taxonomy, Krathwohl and his colleagues (2002) define factual and conceptual knowledge as follows:

> *Factual Knowledge* - The basic elements that students must know to be acquainted with a discipline or solve problems in it.
>
> *Conceptual Knowledge* - The interrelationships among the basic elements within a larger structure that enable them to function together. (p. 214)

The second dimension of the revised taxonomy depicts the structure of cognitive processes (the verbs): remember, understand, apply, analyze, evaluate, and create. The present study is most concerned with the cognitive process *remember* and, within that category, the sub-process of *recalling*. The revised taxonomy defines *remember* as: "Retrieving relevant knowledge from long-term memory" (Krathwohl, 2002: p. 215).

Table 2

Two-dimensional revision of Bloom's taxonomy

The Knowledge Dimension	*1. Remember*	*2. Understand*	*3. Apply*	*4. Analyze*	*5. Evaluate*	*6. Create*
A. Factual Knowledge	X					
B. Conceptual Knowledge	X					
C. Procedural Knowledge						
D. Metacognitive Knowledge						

Note. Adapted from Krathwohl (2002: p. 216).

To summarize, the current study measures two types of educational objectives, according to the revised taxonomy, recall of factual knowledge and recall of conceptual knowledge. Conceptual knowledge is somewhat more complex and more abstract than factual.

Verbatim versus paraphrased/transposed items

In the current study, recall of factual knowledge was measured using a mixture of verbatim and paraphrased/transposed fill-in-the-blank posttest items. In a study by

Clariana and Koul (2002), the authors utilize similar classifications of test items to assess different dimensions of recall. According to fuzzy-trace theory (Brainerd & Reyna, 1990), "…the representation for any particular event is multiply determined, ranging from exact, verbatim representations to fuzzy, gist-like traces…" (Miller & Bjorklund, 1998, p. 185). Verbatim traces include all of the surface details of the episode, including both TBR information and contextual information. Fuzzy, or gist, traces include the semantic meaning, the essential characteristics and relationships implicit in the TBR information. While children rely heavily upon verbatim traces, adults tend to favor the use of gist traces often at the expense of verbatim memory (Brainerd, Reyna, & Brandse, 1995). While the entire verbatim trace is formed at encoding (when the episode is perceived), fuzzy traces arise as the individual processes the information and integrates it into existing schemas or constructs new schemas. While verbatim traces are stronger immediately after encoding, they tend to decay more quickly over time, so adults are able to remember less of the exact *details*. On the other hand, gist traces decay more slowly, so adults tend to remember relatively more *gist* than verbatim information.

In the present study, the activation strength of verbatim traces is assessed through verbatim fill-in-the-blank items. The verbiage used in these verbatim items corresponds directly with that used to describe the key term in the lesson, as demonstrated by the example in Table 3. Note that factual *verbatim* (FV) questions ask for the key term with wording and sentence structure that mimics that of the key term in the reading passage. With factual paraphrased/transposed (FPT) questions, the order of the item's sentence structure is transposed (placing the key term at the end) and some wording is altered while retaining the essential meaning, the gist, of the lesson content.

Table 3

Examples of factual and conceptual posttest items

Knowledge type	Posttest item	Relevant excerpt from reading passage
18 Factual Verbatim (FV)	According to Maslow, _____ needs are the basic biological needs for things such as food, water, and sex.	Physiological needs - basic biological needs for things such as food, water, and sex
18 Factual Paraphrased/ Transposed (FPT)	According to Vroom, it is important that a person feel that performing a task will lead to particular reward. He called this feeling _____.	Instrumentality - the confidence an individual feels that performing the behavior will result in a particular outcome.
8 Conceptual Knowledge (CK)	The distinction between intrinsic and extrinsic factors is most closely associated with _____ Theory.	In his Two-Factor Theory of motivation, Frederick Herzberg (1959) argues that there are two types of factors involved in motivation: extrinsic and intrinsic.

Clariana and Koul (2002) argue convincingly that by altering the wording and structure of these fill-in-the-blank (constructed response) items, it is possible to measure the strength of the learner's gist traces. With paraphrased items, learners cannot rely solely on rote memorization, they must understand the meaning of the question and relate that to the meaning stored in memory. Similarly, with transposed questions, learners can only rely on the meaning of the semantic content in the questions. By combining the two characteristics, paraphrasing and transposing, the FPT questions used in this study should suppress verbatim traces in favor of fuzzy, or gist, traces.

Conceptual knowledge (CK) questions tested recall of more implicit distinctions among the eight theories, not key terms. In structure and wording, CK questions were similar to FPT questions in that both tested recall of concepts using slightly different wording and structure than that found in the reading passage. Fuzzy-trace theory predicts that recall of verbatim, factual knowledge benefits most from color.

In summary, recall of factual knowledge is measured using both verbatim (FV) and paraphrased/transposed (FPT) fill-in-the-blank items. Verbatim questions should encourage learners to make use of verbatim traces. Since verbatim traces are generally stronger than gist traces immediately after encoding, one would expect that learners would perform better on FV items than on FPT items in a test of short-term retrieval. Also, since verbatim traces contain episodic contextual information, such as color, one might expect that incidental color would influence retrieval of verbatim traces more than it would retrieval of gist traces. Because FPT items depend more on gist traces and those traces are relatively weak until additional processing has taken place, one might expect that learners would perform poorly on FPT items in a test of short-term memory and would make little use of color information. In a delayed posttest, the results may be the reverse. In Clariana and Koul's (2002) study, where recall was tested immediately after the learning episode, participants generally performed better on verbatim items than gist items. Also, in that study, feedback was found to have a positive effect on gist recall but not on verbatim recall, suggesting that feedback strengthened gist traces without necessarily strengthening verbatim ones. This is further evidence that additional processing strengthens fuzzy traces and not verbatim ones.

Gender and Color

Brannon (1999) notes that most studies describing differences between male and female participants use the term sex differences. However, some researchers object to the overuse of the term as well as its biological implications. Unger (1979) suggests that the term ‘gender’ be used to refer to traditional, social distinctions between males and females and that the term ‘sex’ be reserved for biological distinctions. Similarly, Gentile (1993) proposes the use of the terms biologically sex-linked, gender-linked, and sex-correlated to distinguish between biological and social differences and between causally linked and correlated differences. For the purposes of the present investigation, the terms gender and gender-linked will be used to refer to social distinctions and any correlated differences that may be attributed to those distinctions, respectively.

Eagly (1987; 1997) and Halpern (1994) suggest that research examining gender differences is valuable and Eagly (1987) goes so far as to suggest that psychological studies should report all gender-related findings, regardless of their theoretical significance, if gender is part of the study design. Others have expressed concern that gender research can only serve to exaggerate perceived differences and to perpetuate stereotypes (see Lott, 1997).

Nevertheless, it stands to reason that color as a contextual variable will have a greater effect on individual learners who are more attuned to color and peripheral elements within their field of vision. Although individual differences in how learners make use of color cues remain largely unexplored, an extrapolation of previous findings (Freedman, 1989; Rogers, 1995; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962; Witkin, Oltman, Raskin, & Karp, 1971) suggests that female students are more attuned to the color context cues and may therefore be more likely to be influenced by them than male students.

Moreover, females may have an inherent advantage over males in constructed response tests (De Mars, 1998, 2000). For example, in an analysis of posttest scores for 295 male and 194 female undergraduates who completed introductory atmospheric and oceanic science courses, Weaver and Raptis (2001) found that female students outperformed male students on constructed response items (short answer questions) while the opposite was true with multiple choice items. These results are consistent with a number of studies examining gender differences in test performance (Bolger & Kellaghan, 1990; Breland, Danos, Kahn, Kubota, & Bonner, 1994; Bridgeman & Lewis, 1994; De Mars, 1998, 2000; Mazzeo, Schmitt, & Bleistein, 1993; Murphy, 1982). These patterns seem to apply mainly to tests that are verbal in nature: the female advantage tends to be smaller in constructed response tests involving mathematics, and the male advantage in multiple choice questions tends to be strongest with mathematics content (Bolger & Kellaghan, 1990; Breland et al., 1994; Murphy, 1982). Based on these results, one would expect female participants to outperform males in the current study, since all of the posttest items are verbal and are in a constructed response format.

So why do females appear to have an advantage on verbal constructed response tests? Are women smarter than men? Sherman (1978) notes that although some claims of a female advantage in general intelligence level exist (Jensen, 1969; Wechsler, 1944), most researchers have come to accept that there is no significant difference in general intelligence between genders (Wechsler, 1955, 1958). "The focus of controversy has switched from the question of general intelligence to questions of differences in specific dispositions, especially spatial and mathematical dispositions, and to some extent the verbal disposition" (Sherman, 1978: p. 38).

Breland et al. (1994) posit that females perform better on constructed response items because of their superior verbal abilities. According to Maccoby and Jacklin (1974), females may have a tentative verbal advantage over males until about age 3, and then the advantage resurfaces at about age 10 or 11. From then on, females consistently outperform males in verbal tasks throughout the high school and college years. Jensen (1969) argues that although female children initially develop verbal abilities faster than male children, the gap narrows as children mature. Nevertheless, it appears that the gap later widens again, and females reach a point where they consistently perform better on verbal ability tests (e.g., vocabulary recognition, abstract reasoning, grammar, and articulation, length of statement, verbosity and verbal fluency, and vocalization) than males, although not by a large margin (Lips, Myers, & Colwill, 1978; Sherman, 1978).

The root cause of the female advantage on constructed response tests is clouded in debate between social constructionists on one side, who argue that gender differences are products of socially constructed roles and stereotypes imposed on men and women (Beall, 1993), and those who hold a biological view that differences in hormonal levels before and after birth are the proximate cause of gender-related cognitive differences (Browne, 2002). Nonetheless, there is little argument that a female advantage on verbal, constructed response tasks exists.

So will color help male or female students more? In a study involving 126 female and 53 male undergraduate students, Dwyer and Moore (2001) found a significant interaction between gender and color (p. 313). The criterion posttest scores (3 multiple-choice/recognition measures and 1 drawing/recall measure) of female students appeared to be more sensitive to the effects of color-coding than those of male students. Other studies also suggest that females are more attuned to color than males (Freedman, 1989; Rogers, 1995).

Although these findings are hardly conclusive, taken as whole, they suggest that women may be more aware of the presence of color than men. Whether or not this increased perception of color is beneficial or detrimental to recall task performance is debatable. In this study, the presence of border colors was peripheral to the TBR content and not the intended focus of learning. Participants were not told that color was useful. Color merely appeared around the border of the screen and changed when the topics changed. Thus, color could have easily become a distraction to participants who were unable to ignore the irrelevant visual stimulus. On the other hand, participants who paid more attention to the color may have been more likely to benefit from the color-coding properties.

Summary of Literature

While color is a pervasive variable in computer-based instruction, it is not yet clear whether color makes an effective recall cue. Tulving's encoding specificity principle (Tulving & Thomson, 1973) predicts that retrieval will be optimized if the same color is present both at encoding and retrieval and suppressed if it is only present at one and not the other. Epstein and Dupree's (1977) interpretation of the generation-recognition model predicts that color cuing at retrieval will have no impact on recall, since learners will have already generated their mental lists of associated concepts and will approach recall tasks using a recognition strategy. Contrary to both theories, color cuing at encoding or retrieval may individually impact recall as advanced organizers (Ausubel, 1968) and as sources of motivation, respectively. The outshining hypothesis (Smith, 1986) predicts that more difficult test items will be more context-dependent than easy items, due to their different activation strength levels. Conversely, fuzzy trace theory (Brainerd & Reyna, 1990) suggests that low-level verbatim traces will be influenced more by context cues than higher level gist memory. The research suggests that there will be gender differences favoring females, due to the constructed response format used in the posttest (Weaver & Raptis, 2001). Also, it is possible that female learners will be more attuned to, and presumably more influenced by, color than male learners. Whether that leads to higher or lower recall task performance is unclear.

Chapter 3

RESEARCH METHOD

The main purpose of this investigation was to examine the relative context effects of border colors presented in a computer-based lesson (at encoding) and/or in a fill-in-the-blank posttest (at retrieval). The posttest was designed to assess recall of factual and conceptual knowledge. The display of border color was manipulated as a local, external context cue. Within this framework, the current investigation was designed to address the following questions:

1. Does the presence of border colors in the lesson or posttest enhance recall task performance?
2. Does *matching* the border color displayed in the posttest with that displayed in the lesson optimize posttest achievement?
3. Is there an interaction between the type of factual recall task assessed (involving verbatim and paraphrased/transposed test items) in the posttest and the context effects of border color?
4. Is there an interaction between posttest item difficulty (easy and difficult) and the context effects of border color?
5. Does recall of factual knowledge vary from that of higher-level conceptual knowledge in terms of the context effects of color?
6. Is there an interaction between gender and the context effects of color?
7. Does *mismatching* the border color displayed in the posttest with that displayed in the lesson undermine posttest achievement?

Participants

Participants were 196 undergraduate students. Initially, 276 students were recruited from a subject pool consisting of about 1,000 students completing a 200-level *Introduction to Psychology* course at a major northeastern research university for general education credit. Of those initial 276, 10 cancelled their appointments; 7 simply failed to attend; 44 were unable to participate due to technical problems related to the laboratory facilities; 12 were unable to complete the experiment; and 7 were dropped from the experiment because of incorrect responses in the color-blindness test. Altogether, 80 participants (about 29% of those recruited) were either unable to participate or participated but yielded unusable data results. This brought the total number of participants to 196.

Students received course credit for participating in the study but not for performance. They were mostly freshmen (63.3%) and sophomores (21.4%) and varied widely in academic majors (14.3% were undecided), as one might expect from students in a typical general education course. Students aged in range from 18 to 35, although 89.5% were under the age of 21. More than half were female (58.2% female, 41.8% male). Mean SAT scores (self-reported) were moderately high (Math: 615/800, Verbal: 587/800).

Students were computer-literate. On a 5-point Likert Scale, 95.9% rated themselves average or better in terms of computer proficiency and 98.5% rated themselves average or better in terms of Internet experience. Most, however, reported never taking an online course of any sort (90.3%).

Students were generally unfamiliar with the instructional content used in the experiment (Theories of Achievement Motivation). Most had never taken an industrial/organizational psychology course (90.3%) or a course on motivation (95.9%). Scores on a test of prior knowledge about industrial/organizational psychology concepts was relatively low (mean score 39.5% on a 17-item multiple-choice test).

Instructional Materials

All of the materials used in the investigation were developed in the form of a computer-based software application using Macromedia's Authorware 5.2, a multimedia-authoring tool. The instructional materials that the participants were asked to examine consisted of three basic elements: the reading passages, the border color, and the main menu screens.

Reading passage

Participants in all treatment groups read a 3,774-word passage (on-screen) dealing with theories of achievement motivation (see appendix A for transcript). This industrial and organizational psychology subject matter was selected because it was believed to be relevant and yet unfamiliar to students in an *Introduction to Psychology* course. This strategy is supported by Morrison, Ross, Gopalakrishnan, and Casey (1995). In their study, involving graduate-level *Instructional Systems Design* and *Educational Psychology* classes, they observed much larger effect sizes when classes were given content that was relevant to their respective fields of study.

The "Theories of Achievement Motivation" content was drawn from the course content materials of a 300-level Industrial and Organizational Psychology course (Peck, 2000) and its text book (Spector, 2000). The topics in the passage were reviewed and approved by an instructor and two research assistants in the Psychology Department of a northeastern research university.

The passage was estimated to be at the 12th grade reading level (per the Flesch-Kincaid rating provided by Microsoft Word). In part due to the technical nature of the subject matter, the passage was rated as difficult (38.3) on the Flesch reading ease scale (a description of the scale is available at http://www.wordinfo.com/how_to/dialogs/Mwdialog00000328.html).

The passage contained primarily text but also included simple illustrations, still photographs, and other graphic elements. Photographs were shown in grayscale mode, as were all of the illustrations, buttons, and other graphic elements on the screen. The only color that appeared on any instructional screen was the experimental treatment being manipulated. All students viewed the same grayscale graphic elements regardless of treatment condition. No animations, audio, or video were used in the lessons. In some places, participants were asked to click a graphic element on the screen to reveal additional text, but generally the amount of interaction was limited.

The reading passage was chunked into eight sections, each describing the main attributes of a theory of achievement motivation:

1. Maslow's Hierarchy of Needs Theory (1954)
2. Alderfer's ERG Theory (1972)
3. Herzberg's Two-factor Theory (1959)
4. Skinner's Reinforcement Theory (1938)

5. Vroom's Expectancy Theory (1964)
6. Bandura's Empowerment Theory (1982, 1986)
7. Adams' Equity Theory (1965)
8. Locke and Latham's Goal-setting Theory (1990).

Each of the eight sections, or lessons, was chunked into four screens containing the following headings: (1) *Overview*, (2) *Major Distinctions*, (3) *Implications*, and (4) *Key Terms*. The headings were standardized for all eight lessons in order to control and minimize the potential context effects of headings (Hartley & Trueman, 1983; Loman & Mayer, 1983; Lorch & Lorch, 1996). Altogether then, the reading passages constituted 32 screens. The names of the theories were shown as headings only on the main menu.

Border Color

The border colors themselves were displayed as thick rectangular borders around the content area of the screen with a swirl that ran down the right side of the screen (see Figure 1 for an example). Table 4 identifies the colors assigned to the eight lesson topics. It also identifies the hexadecimal red-green-blue (RGB#) values for the border colors used. The uncued lessons and posttests contain no screen border element at all. In other words, the space that the border color might have otherwise occupied was simply white.

Table 4

Color designations for each instructional lesson

Lesson Topic	Border Color	RGB Value
Hierarchy of Needs Theory	Blue	#1217EA
ERG Theory	Brown	#8B4F15
Two-factor Theory	Red	#C5321F
Reinforcement Theory	Teal	#206D76
Expectancy Theory	Green	#1CB71C
Empowerment Theory	Periwinkle	#6CD8E5
Equity Theory	Orange	#FCB202
Goal-setting Theory	Purple	#7709BD

Uncued Lesson Screen

Cued Lesson Screen

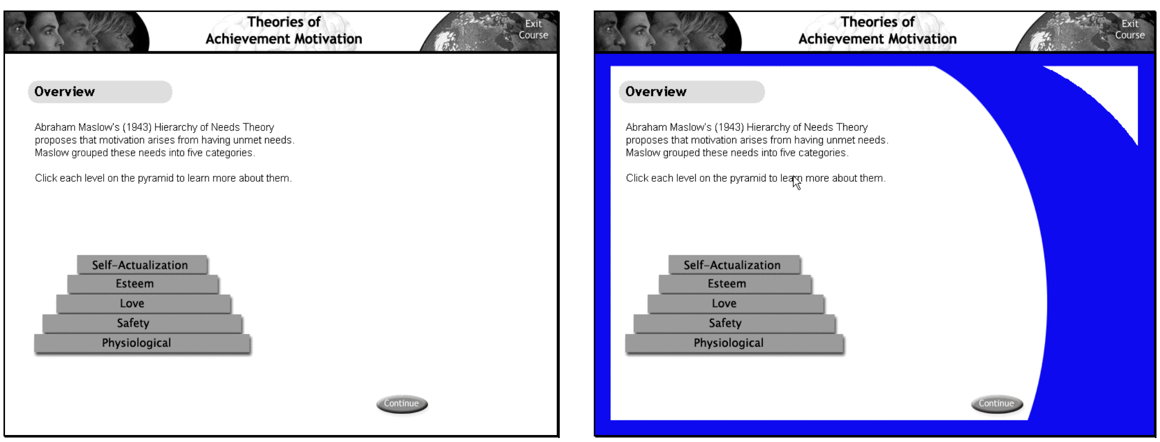

Figure 1. Sample screens demonstrating a typical screen without border colors (on the left) and with border colors (on the right)

Main menu

The main menu (see Figure 2) contained a list of theories from which the subject could select. Next to each theory was a circular button that the subject clicked to begin the lesson. Participants in the No Color lesson conditions had gray buttons, while participants in the Color lesson conditions had color-coded buttons. The topics were made accessible one by one as the participants went through the course. In other words, the second button would not be active until the first lesson was completed, and so on. This controlled for sequence.

In the main menu, and throughout the course, a header section at the top of the screen was visible. In it appeared decorative graphics and the name of the course "Theories of Achievement Motivation." The header also contained an "Exit Course" button to allow participants to terminate the session.

As shown in Figure 2, the main menu had five buttons: one for each of four theories and one button to launch the posttest. However, there were 8 theories covered in all. The 8 reading passages and the posttest were separated into two sets. After participants completed one 4-theory set and the first half of the posttest, they were shown the second menu so that they could complete the second 4-theory set and the second half of the posttest. The sequence of the sets was crossed so that half of the participants completed sequence A-B while the other half completed sequence B-A. This eliminated any sequence and practice effects. Note that in actuality, 49.5% completed sequence A-B and 50.5% completed sequence B-A.

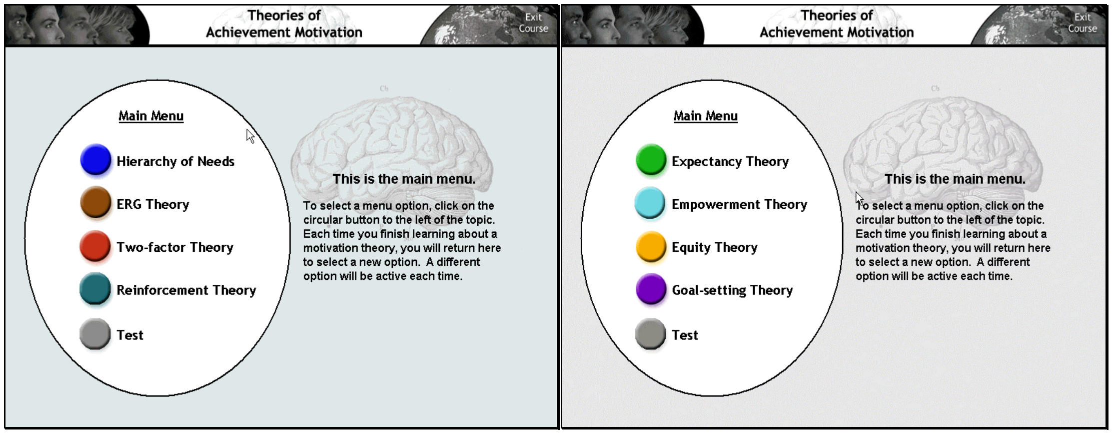

Figure 2. Screen captures of Main Menu A and B

Data Collection Instruments

Pre-Survey

A short computerized survey (see Figure 3) collected information about each subject's age, gender, major, year (i.e., freshman, sophomore, etc.), prior educational experiences related to motivation theories and industrial/organizational psychology, and level of computer experience.

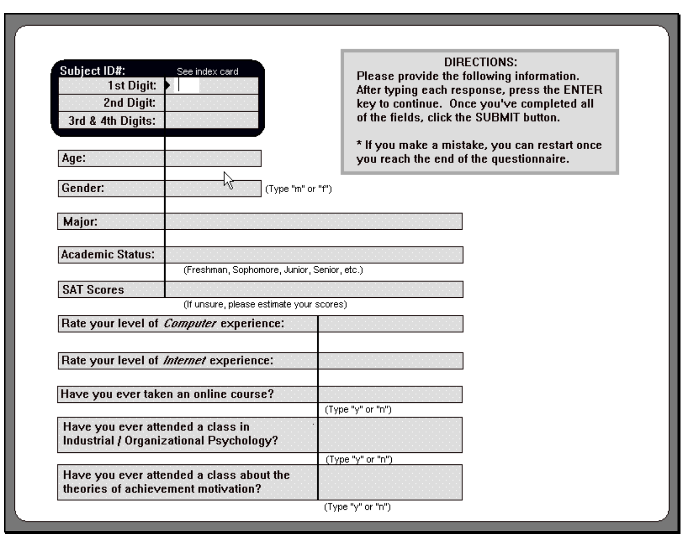

Figure 3. Screen capture of the Pre-Survey

Prior knowledge test

The prior knowledge test consisted of 17 test items, adapted from the test bank (Kravitz & Martin, 2000) of an industrial and organizational psychology text book (Spector, 2000). Psychology Department researchers are currently using the selected test items as a posttest for measuring industrial/organizational psychology knowledge. In a small pilot study, involving 12 Special Education undergraduates, prior knowledge test achievement had a very high correlation with posttest achievement ($r = .937$, $p < .05$). Mean scores for prior knowledge were .30, not significantly better than chance guessing. However, in the current study, prior knowledge test achievement had a very low correlation with overall posttest achievement ($r = .166$, $p < .05$). Even when compared with just the factual knowledge criterion measure, the correlation was low ($r = .182$, $p < .05$). Due to this low correlation, prior knowledge was discarded as a covariate.

Fill-in-the-blank posttest

A 46-item fill-in-the-blank posttest (see Appendix C) was generated from the reading passages. Some test items were adapted from the test bank (Kravitz & Martin, 2000) that accompanied the textbook being used for the reading passage (Spector, 2000), however, most items were created by the researcher.

The 46 test items were used in an earlier *pilot* study. Cronbach's Alpha, the estimate of internal reliability, was .69 and the overall mean percentage score was 27.7%. Two test items were deleted because of low item inter-correlation (item discrimination) values and several others were revised.

The posttest used in the *current* investigation was made up of three main criterion measures: recall of factual knowledge with verbatim (FV) and with paraphrased/transposed (FPT) test item cues as well as recall of conceptual knowledge (CK). The 18 FV test items, the 18 FPT items, and the 8 CK items were split by median difficulty level into easy and hard items, using data from the pilot study. Altogether then, the 44-item posttest was composed of six criterion sub-measures, listed below with their respective item difficulty (*p*) ranges:

- Factual – verbatim - easy (9 items): $.42 \le p \le 1.00$
- Factual – verbatim - hard (9 items): $0 \le p \le .33$
- Factual - paraphrased/transposed - easy (9 items): $.25 \le p \le .75$
- Factual - paraphrased/transposed - hard (9 items): $0 \le p \le .17$
- Conceptual – easy (4 items): $.17 \le p \le .33$
- Conceptual – hard (4 items): $.08 \le p \le .17$

For the full study, Cronbach's Alpha for the entire 44-item posttest (n = 196) was calculated at .85 (see Table 5). The reliability estimates for the factual-verbatim, factual-paraphrased/transposed, and conceptual knowledge portions of the posttest were .74, .70, and .38, respectively. Cronbach's Alpha measures reliability, i.e., a high alpha indicates that the test items are measuring performance consistently. The calculation of Cronbach's Alpha is sensitive to the number of test items used and the test's variance. When the number of test items is high and the posttest variance is high (relative to item variance), the alpha is high. Conversely, when there are few test items in the criterion measure and the posttest variance is low, the alpha is low. The latter scenario explains why the alpha for the conceptual knowledge criterion measure is low. The number of test items was small (4) and the test variance was relatively low.

Table 5

Posttest item difficulty and alpha levels for each criterion measure

		Item Difficulty			
	n	Min	Max	*M*	Alpha
FVE	9	0.49	0.70	0.60	0.68
FVH	9	0.12	0.47	0.27	0.53
FV	18	0.12	0.70	0.44	0.74
FPTE	9	0.19	0.61	0.32	0.58
FPTH	9	0.03	0.19	0.10	0.53
FPT	18	0.03	0.61	0.21	0.70
CKE	4	0.18	0.28	0.22	0.21
CKH	4	0.10	0.19	0.15	0.33
CK	8	0.10	0.28	0.18	0.38
Total	44	0.03	0.70	0.30	0.85

Note. These represent values from the item analysis of the full study posttest, not the pilot study.
FV = Factual knowledge with verbatim cues,
FPT = Factual knowledge with paraphrased/ transposed cues,
CK = Conceptual knowledge

Table 5 shows the range of item difficulty values for each criterion measure, based on data from the full study. Item difficulty is the proportion of participants who answered the item correctly (i.e., the number of participants who answered the question correctly divided by the total number of participants who answered the question). Note that the eight conceptual knowledge posttest items ranged in item difficulty from .10 to .28 (28% answered the easiest item correctly) and the average item difficulty was .18. Item difficulty levels for factual verbatim and factual paraphrased/transposed items ranged from .12 to .70 and .03 to .61, respectively.

Although item difficulty levels for these fill-in-the-blank posttest items were low in comparison to typical multiple-choice recognition test items, they are not unusually low for tests involving recall, or constructed response tasks (Graf & Mandler, 1984). Also, consider that participants were asked to learn a large amount (3,774-word passage) of highly technical, abstract content in a short period of time (one hour time limit) by simply reading the information once and memorizing it without the aid of practice exercises, notes, or mnemonics. In addition, the average item inter-correlation (or item discrimination) value, calculated as the correlation between performance on an item and performance on the entire posttest (Nitko, 1996) was .35, ranging from .13 to .52. The strong, positive item inter-correlations in the posttest items indicate that the posttest functioned properly and the high overall Alpha (.85) indicates that the posttest measured performance reliably. The low item difficulty values reflect the fact that constructed response items are difficult, particularly when compared to multiple-choice test items.

Another factor to consider with regard to the difficulty of the posttest is the manner in which the posttest was scored. Items were scored automatically by the CBI application, and items were scored as *correct* only when a participant entered the specific key term sought. Responses were checked manually and marked *correct* only if they had been marked as *incorrect* due to a participant's spelling error. Sometimes, participants listed terms that were related to the correct answer, perhaps even having similar meanings. While an instructor might have, under typical classroom conditions, offered partial credit or even full credit for those "close" answers, such responses were marked as *incorrect* for the purposes of this investigation. To accept some close answers and not others would have introduced further subjectivity.

The inter-correlations for all six posttest sub-measures were computed and reported in Table 6. Note that the correlation values were low to moderate but significant at the .01 level.

Table 6

Inter-correlations for all posttest sub-measures

	FVE	FVH	FPTE	FPTH	CKE	CKH
FVE	1.00	0.48^{**}	0.53^{**}	0.41^{**}	0.41^{**}	0.32^{**}
FVH	-	1.00	0.56^{**}	0.57^{**}	0.43^{**}	0.26^{**}
FPTE	-	-	1.00	0.56^{**}	0.48^{**}	0.25^{**}
FPTH	-	-	-	1.00	0.40^{**}	0.36^{**}
CKE	-	-	-	-	1.00	0.27^{**}
CKH	-	-	-	-	-	1.00

* $p < .05$, ** $p < .01$

Post Survey

A simple Level 1 evaluation instrument (Kirkpatrick, 1998) was administered once all lessons and posttests had been completed (see Figure 4). This survey gauged overall satisfaction with the instructional materials and gathered qualitative feedback to inform future revisions of the materials. Student satisfaction was gauged with a standard five-point Likert-scale questionnaire (5 being the highest score) that asked about the clarity of the directions, how well key points were presented, and overall student satisfaction with the course. Two open-ended questions were asked, prompting participants to share what they liked and did not like about the educational experience.

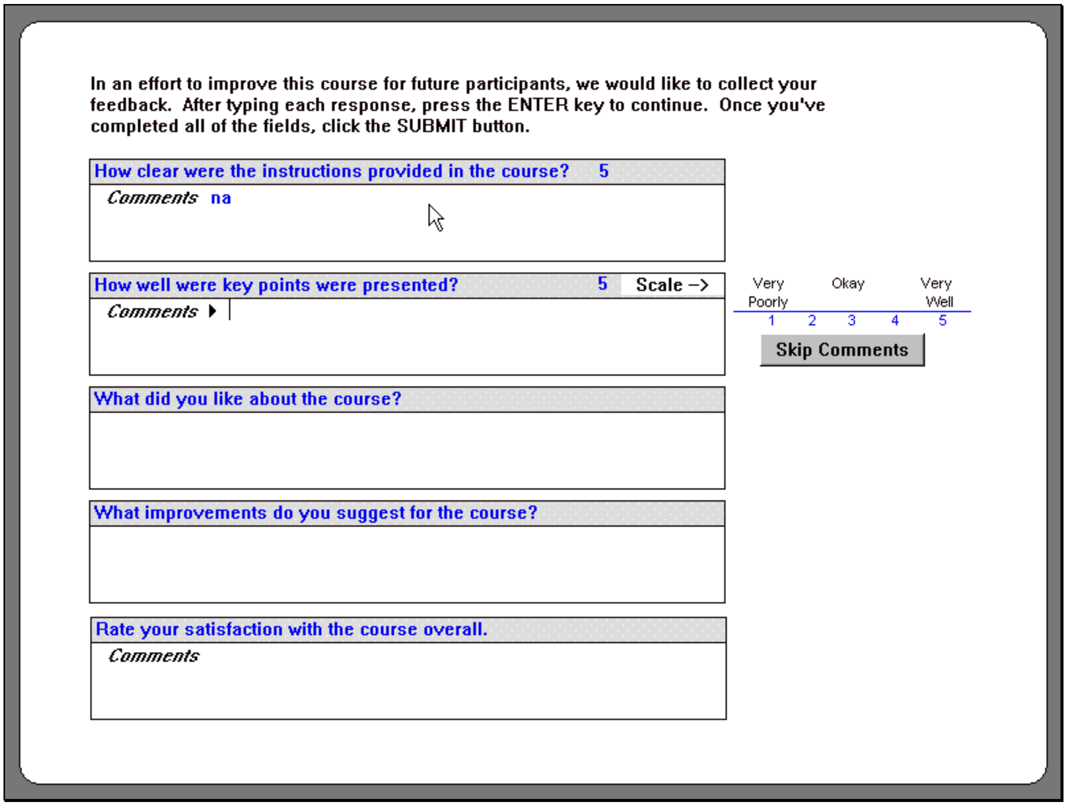

Figure 4. Screen capture of the post survey

When asked to rate the clarity of the instructions, 85% of participants judged it to be high or very high. With regard to how well key points were presented, over half of the participants rated the instructional materials high or very high, and over 93% rated them average or better. Regarding overall satisfaction with the course (see Figure 5), 78% rated it average or better. Student responses to the open-ended questions indicate that many students found the high difficulty of the posttest and their inability to use memorization aids (such as note-taking, practice, review, etc.) frustrating.

Figure 5. Relative frequency distribution of overall student satisfaction ratings

Topic-color matching test

To determine if participants were able to recall the relationship between the color of the borders and the general topic heading assigned to each color (i.e., the theory), a short matching test was administered to students who had completed one of the four color versions of the course (n = 154). Participants were shown a pallet of colors, given the name of a theory, and asked to identify the color that corresponded to the theory by double-clicking on the color, and then another theory would appear, and so on for all eight theories (see Figure 6).

Cronbach's Alpha was .25 and the overall mean score was 1.08 (*SD* = 1.10) out of a maximum score of 8 (or 13.5% correct), not much better than chance guessing (1 out of

8, or 12.5% correct). Since the participants were never told that the colors themselves were important, one would expect that mean scores on this topic-color matching (TCM) test would not exceed the level of chance guessing. These results are described in the exploratory analysis section.

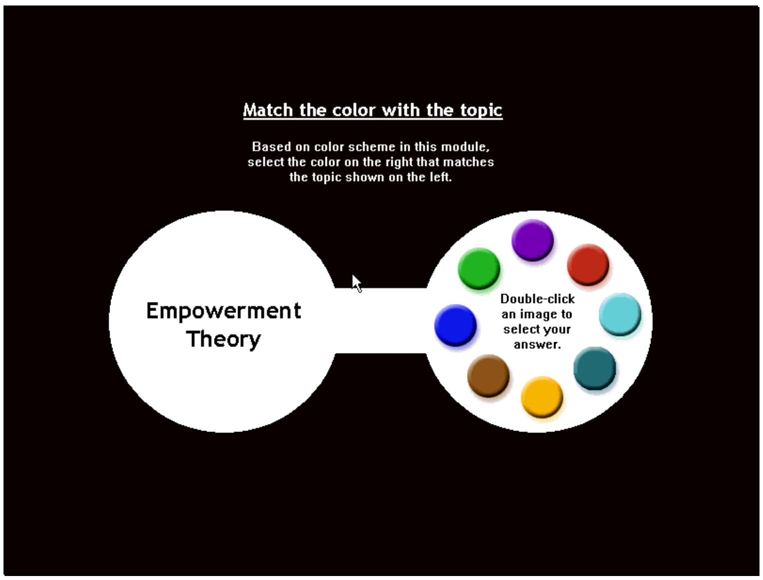

Figure 6. Screen capture of the topic-color matching test

Color blindness test

A color-blindness test — developed by Shinobu Ishihara at the University of Tokyo (available at http://www.copresco.com/links/colblind.htm) and commonly administered by eye doctors – was administered to ensure that all participants completing color versions of the treatments were able to perceive the border colors being manipulated (see Figure 7). The test consisted of 6 circular images, or plates. Each plate contained a number embedded in a field of multi-colored dots. The participants were shown the six images, one at a time, and asked to type the number on the screen. People with normal color vision are able to see the numbers through the dots, but people with red-green color-blindness (the most common form) had difficulty identifying the number correctly.

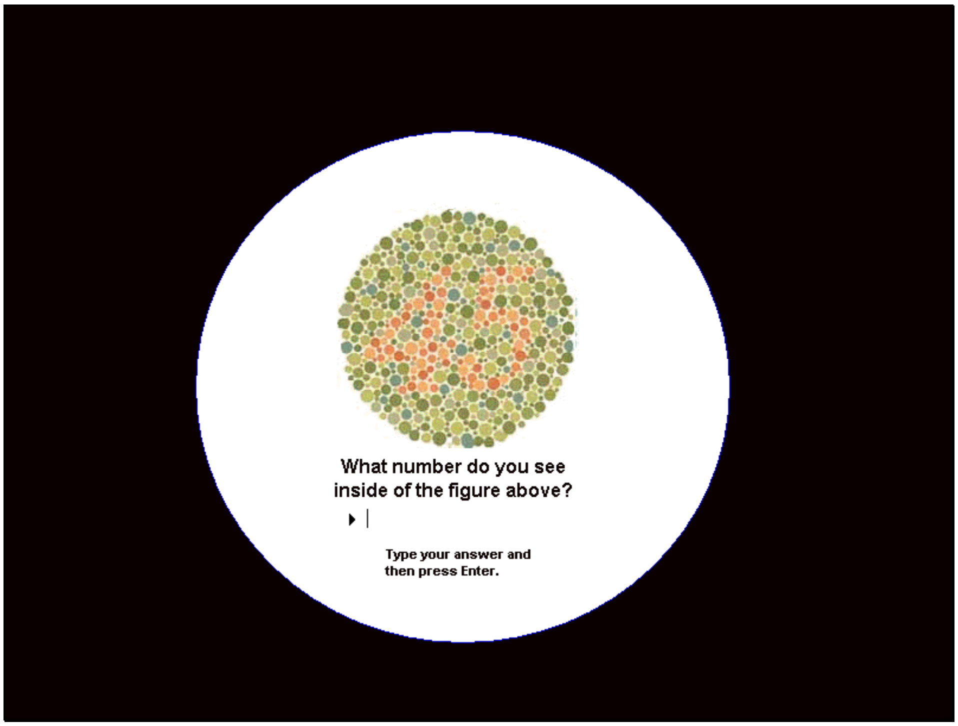

Figure 7. Screen capture of color-blindness test

Changes Resulting from Pilot Study

A preliminary investigation was conducted with a sample of 44 undergraduate Introduction to Psychology students at a major northeastern research university (see Appendix D for pilot study design and results). The main purpose was to determine the reliability of the measurement instruments. Participants were asked to complete a computer-based tutorial with embedded exercises as well as a posttest. In the experimental conditions, the navigation bars on each screen were color-coded to correspond with, or mismatch with, certain topics in the to-be-remembered content. Both recall (fill-in-the-blank) and recognition (multiple choice) were assessed. No significant color context effects were found.

However, the pilot study yielded a several improvement ideas for the main study.

1. The pilot study indicated that there was an interaction, albeit not significant, between setting and posttest achievement. Participants who completed the experiment in large-group sessions situated in a computer lab (M_{lab}=21.17, SD = 3.762, n = 12) performed better on recognition tasks than participants who completed the experiment in an office setting (M_{office} = 20.20, SD = 3.084, n = 10). On constructed response tasks, however, office participants (M_{office}=13.55, SD = 3.024, n = 10) performed better than lab participants (M_{lab}=12.36, SD = 3.146, n = 12). This global context interaction suggested that only one setting should be used in the main study.

2. In the pilot study, lesson headings were displayed on each screen and page headings were displayed in a navigation bar that ran down the left side of the screen. After considering the studies by Epstein and Dupree (1977; 1979) showing strong context effects for category headings, it seemed prudent to remove both types of headings.

3. Although the posttest in the pilot study had a moderately high reliability (.77), the posttest items were not designed to correspond to different educational objectives. Therefore, a number of the questions were re-written, necessitating further piloting of the instruments.

4. The pilot also revealed the possibility that the effects of prior knowledge were distorting context effects. A test of general industrial and organizational psychology prior knowledge was developed so that it could be used as a covariate.

5. In the pilot study, the eight motivation theories were chunked into five categories. The unusually low topic-color matching scores suggested that this approach might have eroded the association between topic and color (since one color might have been associated with two or three theories). Also, the amount of content per lesson varied widely. As a result, the content was re-chunked so that the eight theories were grouped into eight sections of approximately equal size (about 400 words in 4 screens).

6. In the pilot study, participants were given an opportunity to rehearse test items by completing embedded practice exercises in the lesson. It is possible that doing so may have increased the activation strength of the content, thereby reducing its context dependency. In the main study, no practice exercises were used.

Experimental Context

Although steps were taken to ensure that each subject's experience was consistent with those of the other participants, some differences across participants and sessions were likely. During the experiment, there were three primary environmental factors of concern to the researchers: the computer labs, the computer equipment, and the proctors.

There were two computer labs utilized. The maximum number of participants able to register for each time slot was either 17 or 31, depending on the computer lab used. In both cases, the maximum number of participants was half of the maximum seating capacity of the room, which enabled the researchers to space participants two seats apart from one another. A little more than half of the participants participated while in the larger room (53.6%). The smaller room was roughly 5,000 square feet, while the larger room was approximately three times the size. Both rooms were rectangular in shape and the seating was arranged in basically the same format for both rooms, with four rows of tables and chairs that faced away from each other. The main difference was in the length of the rows and the distance between rows.

Due to the larger number of computers that needed to be set-up and the larger number of students that needed to be checked-in, sessions in the larger room generally started about 5 to 10 minutes later than in the smaller room. This may have contributed to some students being unable to finish on time or rushing to get finished.

The monitor is an important consideration, since the technologies used for flat screen monitors, such as LCD (liquid crystal display) and plasma display, are significantly different from the older CRT (cathode-ray tube) technology used in VGA and SVGA monitors. As a result, the intensity of colors and the visual display of a course in general can be different, depending on the type of monitor. All of the monitors used were *Dell* 17-inch flat screen LCD monitors, set to run at an 800 pixel x 600 pixel screen resolution.

The computer-based instructional program was run over a high-speed intranet connection. All of the computers were equipped with a standard mouse and keyboard. The computers were all IBM-compatible desktop PCs (personal computers) manufactured by *Dell*, running Microsoft's *Windows XP Professional*$_{TM}$ operating system. The ones in the larger room were *Dell Optiplex*$_{TM}$ *GX150* desktops with Pentium III, 1.0 GHz processors, and the ones in the smaller room were Dell *Optiplex*$_{TM}$ *GX240* desktops with Pentium IV, 1.8 GHz processors. Despite the difference in processor speed, the run speed of the course did not appear to be affected. This is because the course ran from a single 2 MB Authorware executable file, so the memory load was only an issue for opening the initial screen. After that preliminary hurdle, there was no incremental memory consumption as there would be with typical html-based courses, which load screen by screen. Also, the fast speed of the network connection ensured that the executable file downloaded quickly.

Proctors consisted of 16 graduate students (primarily doctoral students) from the Instructional Systems program of a northeastern university and one undergraduate research assistant. All proctors were given a job aid in advance of the study, which explained the procedures they were to follow. In addition, all proctors were scheduled to arrive 30 minutes before the start of the experiment for an orientation session. Some were assigned to the check-in table and others to the computer lab. Those in the lab helped setup the computers and monitored the participants. Proctors were asked to log any unusual occurrences, including questions asked by participants during the experiment. Although proctors were generally aware of the nature of the investigation, they were not specifically told about the various treatment conditions or, more importantly, how the numbers on the index cards corresponded to the treatment conditions. Only the primary researcher knew what the numbers on the index cards meant, and he was not heavily involved in the check-in procedures or the monitoring of participants. The primary researcher ran the proctor orientation, setup the check-in table and the computers, answered questions for the proctors, and ran the debriefing for students after the experiment.

Procedures

Recruitment

Participants were recruited through a major northeastern university's departmental subject pool, consisting of approximately 1,000 undergraduate students registered for general education credit in the course: *Introduction to Psychology*. Based on established subject pool procedures, the study was advertised on the subject pool's web page. A brief description of the study was posted on the web page as well as time slots. Students in the pool were required to participate as participants in five hours worth of research. This particular study counted as one hour toward that requirement.

To register, students logged onto the subject pool web site, picked a study, and then selected from one of the available time slots. Sessions were scheduled over four days and included morning, mid day, afternoon, and evening time slots in order to recruit a wide variety and a large number of students.

Check-in

Upon arrival, participants were asked to read and sign the informed consent form (see Appendix E) in compliance with human participants internal review board (HSIRB) guidelines. Per standard subject pool procedures, the participants' names were written on a log, and then they were given participation receipts. In addition, each subject was given an index card containing the subject's identification number. This number not only identified each subject but also indicated to the computer program which treatment condition and sequence to implement for the subject. For example, subject 1253 was in treatment group 1 (1253), which viewed no border colors in the lesson and no border colors in the posttest. This subject was also in sequence group 2 (1253), which meant that the order of the lesson topics would be lessons 5-8 followed by lessons 1-4. The treatment conditions and sequence conditions were randomized within blocks. The number 53 (1253) was a unique identifier within the block.

Randomization. Index cards were assigned as participants went through the check-in process. However, the index cards were randomly ordered ahead of time. Randomization was accomplished by rolling a standard six-sided die (ignoring the number six) to assign treatment conditions to a list of two-digit subject identification numbers. The two-digit numbers were unique to each subject. A second die was used to assign sequence condition (even-odd numbers). The 4-digit numbers were blocked into sets of ten, so that each set had at least one of the ten possible combinations (5 treatments x 2 sequences). Thirty 10-card sets of index cards were created, and then the thirty sets were randomly ordered.

Getting situated

After participants were checked-in and given their index cards, a proctor led them to their workstations, asked them to sit at the computer, read the directions, and begin when ready. Students were generally seated from the back of the room to the front of the room and were seated two spaces apart whenever possible.

Begin the experiment

On each subject's computer, there was initially an instructions screen (see Figure 8). It explained what the participants would be expected to complete. They were instructed to click the *Continue* button to begin.

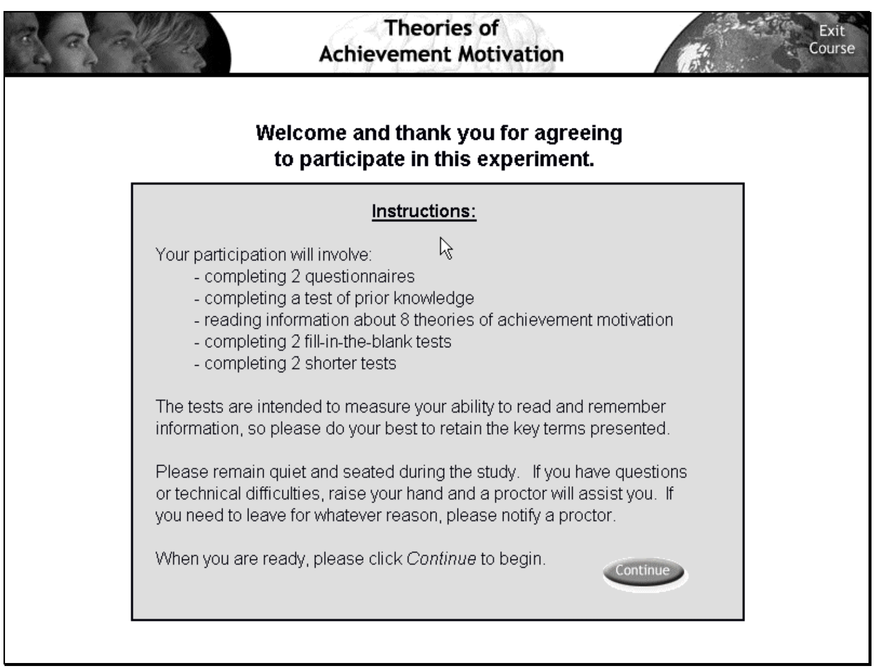

Figure 8. Screen capture of the study introduction

If they had questions, participants were encouraged to raise their hands. Proctors were instructed to answer only general procedural questions (such as “May I start right away? May I use the rest room?”), not questions related to the CBI course. Such questions were to be answered with the response: “Just do your best to interpret the instructions.”

Complete pre survey and prior knowledge test

Upon clicking the *Continue* button, the start date and time were recorded by the program. Next, participants were asked to complete the questionnaire. After that, participants completed the 17-item test of prior knowledge. The computer program automatically recorded all of the data when the subject completed the test.

Complete lessons and posttests

Once participants completed the pre-survey and prior knowledge test, they were branched to the course introduction screen (see Figure 9), which briefly introduced the course topic: Theories of Achievement Motivation.

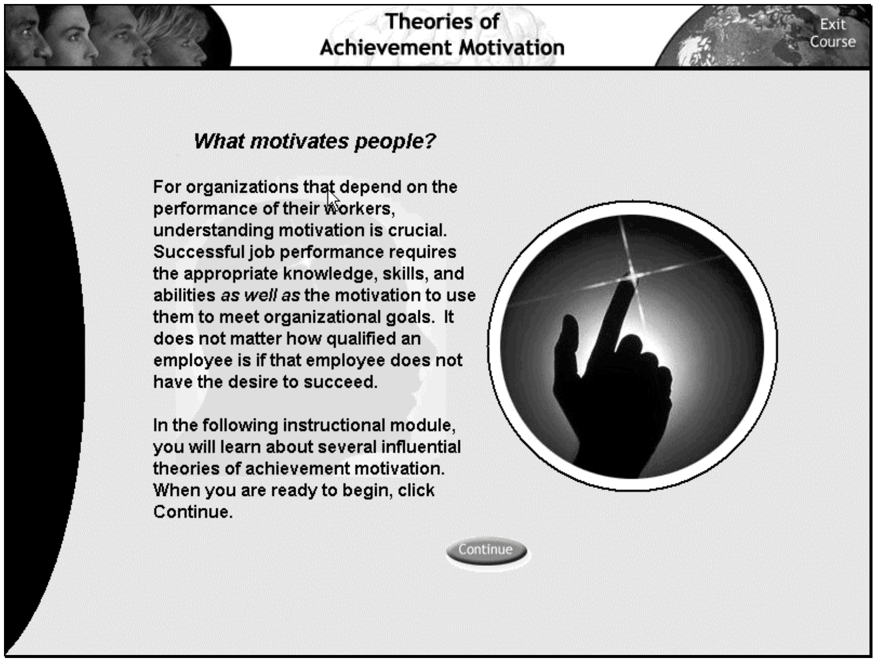

Figure 9. Screen capture of the course introduction

From this screen, participants branched to one of the two main menus (which menu depended on which sequence condition participants were assigned). Although there were eight theories, the main menu only displayed four of them at a time. In order to reduce the cognitive load and prevent fatigue, participants viewed four theories, completed the posttest items related to those four theories, then viewed the other four theories, and completed the remaining posttest items. To combat practice effects, the order of two 4-theory sections was reversed for half of the participants, counter-balancing the effects. In other words, half of the participants were randomly assigned to a sequence in which theories 1 through 4 were presented first and 5 through 8 were presented second, while the other half saw theories 5 through 8 first and theories 1 through 4 last.

Regardless of the sequence, all participants viewed the eight theories and completed the 44-item posttest. It was here, however, where participants experienced one of five different cuing conditions (C = color, NoC = no color, MisC = mismatched color hue):

1. NoC x NoC: This was the control condition in which no border color was presented in the lesson or in the posttest.
2. NoC x C: No border colors were displayed in the lesson, but border colors were displayed in the posttest.
3. C x NoC: Border colors were displayed in the lesson, but they were not displayed in the posttest.
4. C x C: Border colors were displayed in the lesson, and corresponding border colors were displayed in the posttest.
5. C x MisC: Border colors were displayed in the lesson, and intentionally mismatched colors were displayed in the posttest.

Complete the post survey

Once participants completed all of the lessons and posttests, they were asked to provide feedback about the learning experience. The questions were asked one at a time. Participants typed a Likert scale value (1 through 5) and were then given the opportunity to enter comments (which they could skip). This was repeated for the five questions.

Complete the topic-color matching test

After the post survey, participants were branched to the topic-color matching test. This test was only administered to participants in the four experimental conditions (NoC x C, C x NoC, C x C, and C x MisC), not to those in the control condition (NoC x NoC). After the last theory, the participants were asked to select their favorite color using the same color palette.

Complete the color-blindness test

After completing the topic-color matching test, participants branched directly to the Ishihara color-blindness test. The test was given after the course was completed to avoid cuing the learner in on the importance of color. After the test was administered, a results page displayed all six images, their correct answers, and the responses that the participants entered (see Figure 10). Participants were instructed: "… if you were not able to identify the numbers in these figures, you may wish to consult with an ophthalmologist."

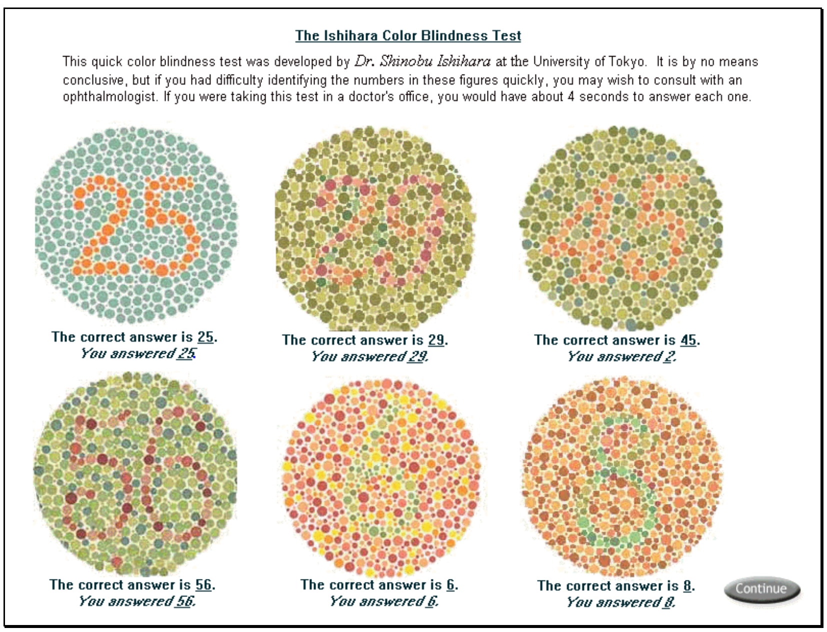

Figure 10. Screen capture of color-blindness test results

Debriefing

After the color-blindness test was completed, participants branched back to the main menu where they were asked to write the word "Finudo" (Italian for "finished") on their index cards and hand them to the proctor waiting outside. The primary researcher was waiting outside in order to debrief students as they finished. He checked to make sure that the word "Finudo" was written on the index card (this let the researcher know that the subject had reached the final screen). On their way out, participants were given a debriefing document (see Appendix F) that explained the study and were given the opportunity to ask questions. The researcher summarized the intent of the study and solicited comments.

Collect data

As participants completed the course and left the computer lab, a proctor clicked the Exit Course button and logged off the computers. Upon exiting, the program automatically created a back-up copy of the data. Both data files were automatically saved to the network and were later downloaded onto a zip disk.

Post data collection

Subject pool procedures required that the primary researcher log onto the pool's web site and indicate which students had attended the experiment, which failed to show, and which rescheduled. Through this process, students were given credit for their participation in the study.

Design

A posttest only control group design (Gall, Borg, & Gall, 1996) with an additional outside control group was used for this experimental study (see Table 7). There were 3-between and 2-within factors. Lesson condition (no border colors or with border colors in the lesson) and posttest condition (no border colors or with border colors in the posttest) were the primary independent variables. The external control group viewed border colors in the lesson and *mismatching* border colors in the posttest. Note that the external control (C x MisC) was included in order to test Hypothesis 7 (that mismatching the border colors in the posttest would interfere with recall).

Gender was included as a third between-subjects variable in order to test Hypothesis 6. The type of posttest item cues (factual verbatim and factual paraphrased/transposed), item difficulty (high and low), and knowledge type (factual and conceptual) were used as within-subjects variables in order to test Hypotheses 3, 4, and 5.

The primary dependent variable for all research questions was recall of factual and conceptual knowledge as measured by achievement on a 44-item fill-in-the-blank posttest described earlier. This posttest measure was split into six sub-measures based on item type and difficulty.

Table 7

Depiction of the completely crossed design with an external control

Lesson \ Posttest	With No Border Colors	With Border Colors	
With No Border Colors	NoC x NoC	NoC x C	
With Border Colors	C x NoC	C x C	* C x MisC

* External control group in which participants view border colors in the lesson and mismatching border colors in the posttest

Data Analysis

The data for each subject was pulled from the individual data files and compiled into one master data set. Although the program scored all of the tests, and the program was given alternate spellings to use for scoring, a manual inspection of the fill-in responses was performed to correct for any unforeseen misspellings.

Color-blindness results were reviewed immediately to determine if any participants needed to be dropped from the study. The statistical analysis for each hypothesis proceeded as follows:

In order to control the inflation of Type I errors (Tabachnick & Fidell, 1996, p. 376) and to examine the multivariate interactions of interest to this investigation, the analyses for hypotheses one through six were combined into one analysis: a 2 x 2 x 2 x 3 x 2 mixed MANOVA (Kachigan, 1991; Kennedy & Bush, 1985) with repeated measures (*lesson condition*: no color and color x *posttest condition*: no color and color x *gender* x *item type*: factual-verbatim, factual-paraphrased/transposed, and conceptual knowledge x *item difficulty*: easy and hard). The first three were between-subject variables and the last two were within-subjects variables. The dependent measure was recall, which consisted of six sub-measures:

1. Factual-verbatim-easy (FVE)
2. Factual-verbatim-hard (FVH)
3. Factual-paraphrased/transposed-easy (FPTE)
4. Factual-paraphrased/transposed-hard (FPTH)
5. Conceptual knowledge-easy (CKE)
6. Conceptual knowledge-hard (CKH)

In a review of educational research analyses, Keselman and his colleagues (Keselman, Huberty, Lix, & Olejnik, 1998) noted that in 84% of the studies involving MANOVA, researchers failed to interpret the multivariate effects, the interactions of grouping variables and outcome variables, and instead interpreted only the multiple univariate analyses. They caution: "Do not conduct a MANOVA unless it is the multivariate effects that are of substantial interest" (p. 360).

Keselman et al. (1998) also note that repeated measures designs are efficient, insofar as they do not require independent groups of participants for each possible treatment combination, and appropriate for detecting within-subjects treatment effects, in cases where *all* participants are being exposed to *all* levels of a variable (p. 362). In the current study, all participants are exposed to all levels of item type and item difficulty and the interactions of those within-subjects variables with the between-subjects variables are of great interest.

For hypothesis 7, a 2 x 2 ANOVA was used (treatment group: color-matching and color-mismatching groups x gender), with both as between-subjects independent variables and total recall as the dependent variable.

All inferential analysis was conducted using SPSS 11.0 for Windows. The alpha level was set at .05. When using analysis of variance, three analytical and statistical assumptions must be met: normality of distribution, independence of observations, and variance homogeneity (Keselman et al., 1998). Unless otherwise noted, all reported MANOVA and ANOVA results met these assumptions.

Exploratory Analysis

Additional exploratory analysis was also conducted:

1. Examination of the possible interaction between posttest scores (by treatment group) and topic-color matching scores.

2. Examination of the possible interaction between lesson duration and lesson color.

Chapter 4

RESULTS

The posttest means and effect sizes were calculated for each measure and sub-measure by gender. To facilitate the analysis of how border colors influenced posttest scores, the descriptive statistics shown in Tables 8, 9, and 10 dissect the scores by treatment group, by lesson condition, and by posttest condition, respectively. Table 8 suggests that the various experimental treatment conditions influenced posttest scores, but the color-matching group (C x C) did not, as predicted, outperform the other groups. Table 9 indicates that presenting border colors during the lesson (at encoding) consistently enhanced posttest performance. Table 10 indicates that males benefited from border color in the posttest, whether there was color in the lesson or not, while females consistently did not benefit and were often negatively affected by border colors in the posttest. These descriptive findings were assessed through inferential statistics, consisting of primarily analyses of variance. A 3-between, 2-within mixed MANOVA was used to consider hypotheses 1 through 6, while a 2-between-subjects ANOVA was used to consider hypothesis 7. Additional post hoc analysis was used to examine significant effects more closely.

Table 8 *Means and effect sizes for all recall measures by treatment condition and gender*

Lesson x Posttest Condition	*n*	Factual Knowledge: Verbatim Items				Factual Knowledge: Paraphrased/Transposed Items				Conceptual Knowledge					
		Easy		Hard		Easy		Hard		Easy		Hard		Total	
		M *SD*	*ES*	*M* *SD*	*ES*	*M* *SD*	*ES*	*M* *SD)*	*ES*	*M* *SD*	*ES*	*M* *SD*	*ES*	*M* *SD*	*ES*
No Color x No Color															
Male	19	4.26	-	1.79	-	2.37	-	0.42	-	0.53	-	0.32	-	9.68	-
		1.727		1.873		1.770		0.838		0.697		0.671		5.313	
Female	23	5.26	-	3.04	-	3.48	-	1.13	-	1.17	-	0.70	-	14.78	-
		2.115		2.099		2.213		1.217		1.114		0.974		7.710	
Total	42	4.81	-	2.48	-	2.98	-	0.81	-	0.88	-	0.52	-	12.48	-
		1.991		2.075		2.078		1.110		0.993		0.862		7.134	
No Color x Color															
Male	16	6.38	1.22*	2.63	0.45	3.31	0.53	0.75	0.39	0.81	0.41	0.44	0.18	14.31	0.87
		2.094		1.310		1.740		0.931		0.750		0.629		5.121	
Female	24	5.58	0.15	1.88	-0.56	2.38	-0.50	0.88	-0.21	0.71	-0.42	0.50	-0.20	11.92	-0.37
		2.185		1.597		1.765		1.454		0.859		0.780		6.192	
Total	40	5.90	0.55	2.18	-0.15	2.75	-0.11	0.83	0.01	0.75	-0.13	0.48	-0.06	12.88	0.06
		2.158		1.517		1.794		1.259		0.809		0.716		5.841	
Color x No Color															
Male	14	5.00	0.43	2.43	0.34	2.86	0.28	0.79	0.44	0.93	0.58	0.43	0.17	12.43	0.52
		2.909		1.989		1.834		0.893		0.997		0.756		7.251	
Female	26	6.38	0.53	3.08	0.02	3.23	-0.11	1.31	0.15	1.15	-0.02	0.69	0.00	15.85	0.14
		2.118		1.787		1.925		1.644		0.834		0.928		7.170	
Total	40	5.90	0.55	2.85	0.18	3.10	0.06	1.13	0.28	1.08	0.20	0.60	0.09	14.65	0.30
		2.479		1.861		1.878		1.436		0.888		0.871		7.294	

Color x Color																
Male	17	5.41	0.67	2.41	0.33	2.94	0.32	0.94	0.62	0.94	0.60	0.59	0.41	13.24	0.67
		2.123		1.698		1.853		1.478		0.966		0.795		5.901	
Female	20	5.30	0.02	2.55	-0.24	2.85	-0.28	1.15	0.02	1.05	-0.11	0.75	0.06	13.65	-0.15
		2.716		1.605		2.033		1.309		0.945		0.786		7.028	
Total	37	5.35	0.27	2.49	0.00	2.89	-0.04	1.05	0.22	1.00	0.12	0.68	0.18	13.46	0.14
		2.429		1.627		1.926		1.373		0.943		0.784		6.449	
Total															
Male	66	5.23	0.56	2.29	0.27	2.85	0.27	0.71	0.34	0.79	0.38	0.44	0.19	12.30	0.49
		2.292		1.726		1.791		1.064		0.851		0.704		6.015	
Female	93	5.67	0.19	2.65	-0.19	2.99	-0.22	1.12	-0.01	1.02	-0.14	0.66	-0.04	14.10	-0.09
		2.286		1.828		1.997		1.413		0.944		0.866		7.088	
Total	159	5.48	0.34	2.50	0.01	2.93	-0.02	0.95	0.13	0.92	0.04	0.57	0.05	13.35	0.12
		2.292		1.789		1.910		1.292		0.911		0.808		6.703	

Note. The maximum overall posttest score = 44. This breaks down as follows: the factual knowledge criterion measure consisted of 18 verbatim items (9 easy / 9 hard) and 18 paraphrased/transposed items (9 easy / 9 hard); the conceptual knowledge criterion measure consisted of 8 items (4 easy / 4 hard).

Table 9 *Mean scores and effect sizes for all recall measures by lesson condition and gender*

Lesson Condition		Factual Knowledge: Verbatim Items						Factual Knowledge: Paraphrased/Transposed Items						Conceptual Knowledge							
		Easy		Hard		Total FV		Easy		Hard		Total FPT		Easy		Hard		Total CK		Total	
		M	*ES*	*M*	*ES*	*M*	*ES*	*M*	*ES*	*M*	*ES*	*M*	*ES*	*M*	*ES*	*M*	*ES*	*M*	*ES*	*M*	*ES*
	n	*SD*		*SD*		*SD*		*SD*		*SD*		*SD*		*SD*		*SD*		*SD*		*SD*	
No Border Colors in the Lesson																					
Male	35	5.23	-	2.17	-	7.40	-	2.80	-	0.57	-	0.57	-	0.66	-	0.37	-	1.03	-	11.80	-
		2.157		1.671		3.136		1.795		0.884		0.884		0.725		0.646		1.071		5.656	
Female	47	5.43	-	2.45	-	7.87	-	2.91	-	1.00	-	1.00	-	0.94	-	0.60	-	1.53	-	13.32	-
		2.134		1.932		3.555		2.052		1.335		1.335		1.009		0.876		1.487		7.050	
Total	82	5.34	-	2.33	-	7.67	-	2.87	-	0.82	-	0.82	-	0.82	-	0.50	-	1.32	-	12.67	-
		2.133		1.819		3.370		1.936		1.177		1.177		0.904		0.790		1.341		6.498	
Border Colors Present in the Lesson																					
Male	31	5.23	0.00	2.42	0.15	7.65	0.08	2.90	0.06	0.87	0.34	0.87	0.34	0.94	0.39	0.52	0.23	1.45	0.40	12.87	0.19
		2.473		1.803		3.656		1.814		1.231		1.231		0.964		0.769		1.410		6.443	
Female	46	5.91	0.22	2.85	0.21	8.76	0.25	3.07	0.08	1.24	0.18	1.24	0.18	1.11	0.17	0.72	0.14	1.83	0.20	14.89	0.22
		2.430		1.712		3.640		1.960		1.493		1.493		0.875		0.861		1.371		7.116	
Total	77	5.64	0.14	2.68	0.19	8.31	0.19	3.00	0.07	1.09	0.23	1.09	0.23	1.04	0.24	0.64	0.18	1.68	0.27	14.08	0.22
		2.454		1.751		3.664		1.892		1.397		1.397		0.910		0.826		1.390		6.882	

Total																					
Male	66	5.23	-	2.29	-	7.52	-	2.85	-	0.71	-	0.71	-	0.79	-	0.44	-	1.23	-	12.30	-
		2.292		1.726		3.366		1.791		1.064		1.064		0.851		0.704		1.250		6.015	
Female	93	5.67	-	2.65	-	8.31	-	2.99	-	1.12	-	1.12	-	1.02	-	0.66	-	1.68	-	14.10	-
		2.286		1.828		3.605		1.997		1.413		1.413		0.944		0.866		1.431		7.088	
Total	159	5.48	-	2.50	-	7.98	-	2.93	-	0.95	-	0.95	-	0.92	-	0.57	-	1.49	-	13.35	-
		2.292		1.789		3.519		1.910		1.292		1.292		0.911		0.808		1.373		6.703	

Note. The maximum overall posttest score = 44. This breaks down as follows: the factual knowledge criterion measure consisted of 18 verbatim items (9 easy / 9 hard) and 18 paraphrased/transposed items (9 easy / 9 hard); the conceptual knowledge criterion measure consisted of 8 items (4 easy / 4 hard).

Table 10

Mean scores and effect sizes for all recall measures by posttest condition and gender

Posttest Condition		Factual Knowledge: Verbatim Items						Factual Knowledge: Paraphrased/Transposed Items						Conceptual Knowledge							
		Easy		Hard		Total FV		Easy		Hard		Total FPT		Easy		Hard		Total CK		Total	
	n	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*	*M* / *SD*	*ES*
No Border Colors in the Posttest																					
Male	33	4.58	-	2.06	-	6.64	-	2.58	-	0.58	-	0.58	-	0.70	-	0.36	-	1.06	-	10.85	-
		2.292		1.919		3.604		1.786		0.867		0.867		0.847		0.699		1.273		6.256	
Female	49	5.86	-	3.06	-	8.92	-	3.35	-	1.22	-	1.22	-	1.16	-	0.69	-	1.86	-	15.35	-
		2.170		1.919		3.633		2.047		1.447		1.447		0.965		0.940		1.594		7.370	
Total	82	5.34	-	2.66	-	8.00	-	3.04	-	0.96	-	0.96	-	0.98	-	0.56	-	1.54	-	13.54	-
		2.294		1.970		3.771		1.972		1.281		1.281		0.942		0.862		1.517		7.251	
Border Colors Present in the Posttest																					
Male	33	5.88	0.57*	2.52	0.24	8.39	0.49*	3.12	0.30	0.85	0.31	0.85	0.31	0.88	0.21	0.52	0.23	1.39	0.26	13.76	0.47*
		2.132		1.503		2.904		1.781		1.228		1.228		0.857		0.712		1.223		5.477	
Female	44	5.45	-0.19	2.18	-0.46*	7.64	-0.35	2.59	-0.37	1.00	-0.15	1.00	-0.16	0.86	-0.31	0.61	-0.09	1.48	-0.24	12.70	-0.36
		2.416		1.618		3.491		1.884		1.381		1.381		0.905		0.784		1.210		6.565	
Total	77	5.64	0.13	2.32	-0.17	7.96	-0.01	2.82	-0.11	0.94	-0.02	0.94	-0.02	0.87	-0.12	0.57	0.01	1.44	-0.06	13.16	-0.05
		2.294		1.568		3.254		1.848		1.311		1.311		0.879		0.751		1.208		6.107	

Total											
Male	66	5.23 -	2.29 -	7.52 -	2.85 -	0.71 -	0.71 -	0.79 -	0.44 -	1.23 -	12.30 -
		2.292	1.726	3.366	1.791	1.064	1.064	0.851	0.704	1.250	6.015
Female	93	5.67 -	2.65 -	8.31 -	2.99 -	1.12 -	1.12 -	1.02 -	0.66 -	1.68 -	14.10 -
		2.286	1.828	3.605	1.997	1.413	1.413	0.944	0.866	1.431	7.088
Total	159	5.48 -	2.50 -	7.98 -	2.93 -	0.95 -	0.95 -	0.92 -	0.57 -	1.49 -	13.35 -
		2.292	1.789	3.519	1.910	1.292	1.292	0.911	0.808	1.373	6.703

Note. The maximum overall posttest score = 44. This breaks down as follows: the factual knowledge criterion measure consisted of 18 verbatim items (9 easy / 9 hard) and 18 paraphrased/transposed items (9 easy / 9 hard); the conceptual knowledge criterion measure consisted of 8 items (4 easy / 4 hard).

* $p < .05$, ** $p < .01$

Hypotheses 1 through 6

A 2 x 2 x 2 x 3 x 2 multivariate analysis of variance (MANOVA) with repeated measures was used, with *lesson condition* (no color and with color), *posttest condition* (no color and with color), *gender*, *item type* (FV, FPT, and CK items), and *item difficulty* (easy and hard items). The first three were between-subjects variables, and the last two were within-subjects repeated measures. The dependent variable was recall as evidenced by achievement on a fill-in-the-blank posttest.

The results of the 2 x 2 x 2 x 3 x 2 MANOVA with repeated measures (see Table 11) revealed a significant between-subjects interaction between gender and the presence of border colors in the posttest, $F(1, 151) = 6.057$, $p = .015$, with males benefiting from posttest color and the opposite effect for females. No other between-subjects effects reached significance.

As shown in Figure 11, both male (*ES* = .19) and female (*ES* = .22) participants benefited from the presence of border colors in the lesson, although this difference was not significant. However, they were influenced *differentially* by the presence of border colors in the posttest, and this interaction *was* significant. As shown in Table 10, male participants recalled approximately 25% of the TBR content when border colors were *not* present in the posttest (*M* = 10.85, SD = 6.256, n = 33), but recalled 31% of the TBR content when border colors *were* present in the posttest (*M* = 13.76, SD = 5.477, n = 33). The positive effect of border colors in the posttest for males (*ES* = .47) was statistically significant at the .05 level. Conversely, female participants recalled about 35% of TBR content when border colors were *not* present in the posttest (*M* = 15.35, SD = 7.370, n = 49), but less than 29% when border colors *were* present (*M* = 12.70, SD = 6.565, n = 44). This negative effect was substantial (*ES* = -.36) but not statistically significant.

Table 11

2 x 2 x 2 x 3 x 2 MANOVA with repeated measures for border colors and gender

Source	*SS*	*df*	*MS*	*F*	*p*
Between subjects					
Lesson Condition (L)	7.927	1	7.927	1.095	.297
Posttest Condition (P)	.055	1	.055	.008	.930
Gender (G)	16.987	1	16.987	2.348	.128
L x P	3.951	1	3.951	.546	.461
L x G	.508	1	.508	.070	.791
P x G	43.831	1	43.831	6.057*	.015
L x P x G	8.024	1	8.024	1.109	.294
Error	1092.619	151	7.236		
Within subjects					
Item Type (T)	1629.800	1.889	862.759	442.988***	.000
T x L	.121	1.889	.064	.033	.962
T x P	1.075	1.889	.569	.292	.734
T x G	.671	1.889	.355	.182	.821
T x L x P	9.412	1.889	4.982	2.558	.082
T x L x G	1.726	1.889	.914	.469	.615
T x P x G	12.163	1.889	6.439	3.306*	.041
T x L x P x G	.751	1.889	.397	.204	.803
Item Difficulty (D)	725.860	1	725.860	441.500***	.000
D x L	.281	1	.281	.171	.680
D x P	1.145	1	1.145	.697	.405
D x G	.331	1	.331	.201	.654
D x L x P	3.975	1	3.975	2.418	.122
D x L x G	.258	1	.258	.157	.693
D x P x G	3.197	1	3.197	1.945	.165
D x L x P x G	.112	1	.112	.068	.795
Item Type x Item Difficulty (T x D)	266.316	1.753	151.922	110.699***	.000
T x D x L	.706	1.753	.403	.294	.717
T x D x P	7.696	1.753	4.390	3.199*	.049
T x D x G	1.183	1.753	.675	.492	.587
T x D x L x P	6.089	1.753	3.474	2.531	.089
T x D x L x G	.575	1.753	.328	.239	.758
T x D x P x G	.865	1.753	.493	.359	.670
T x D x L x P x G	3.872	1.753	2.209	1.609	.205
Error (T)	555.545	285.248	1.948		
Error (D)	248.256	151	1.644		
Error (T x D)	363.270	264.699	1.372		

* $p < .05$, ** $p < .01$, *** $p < .001$

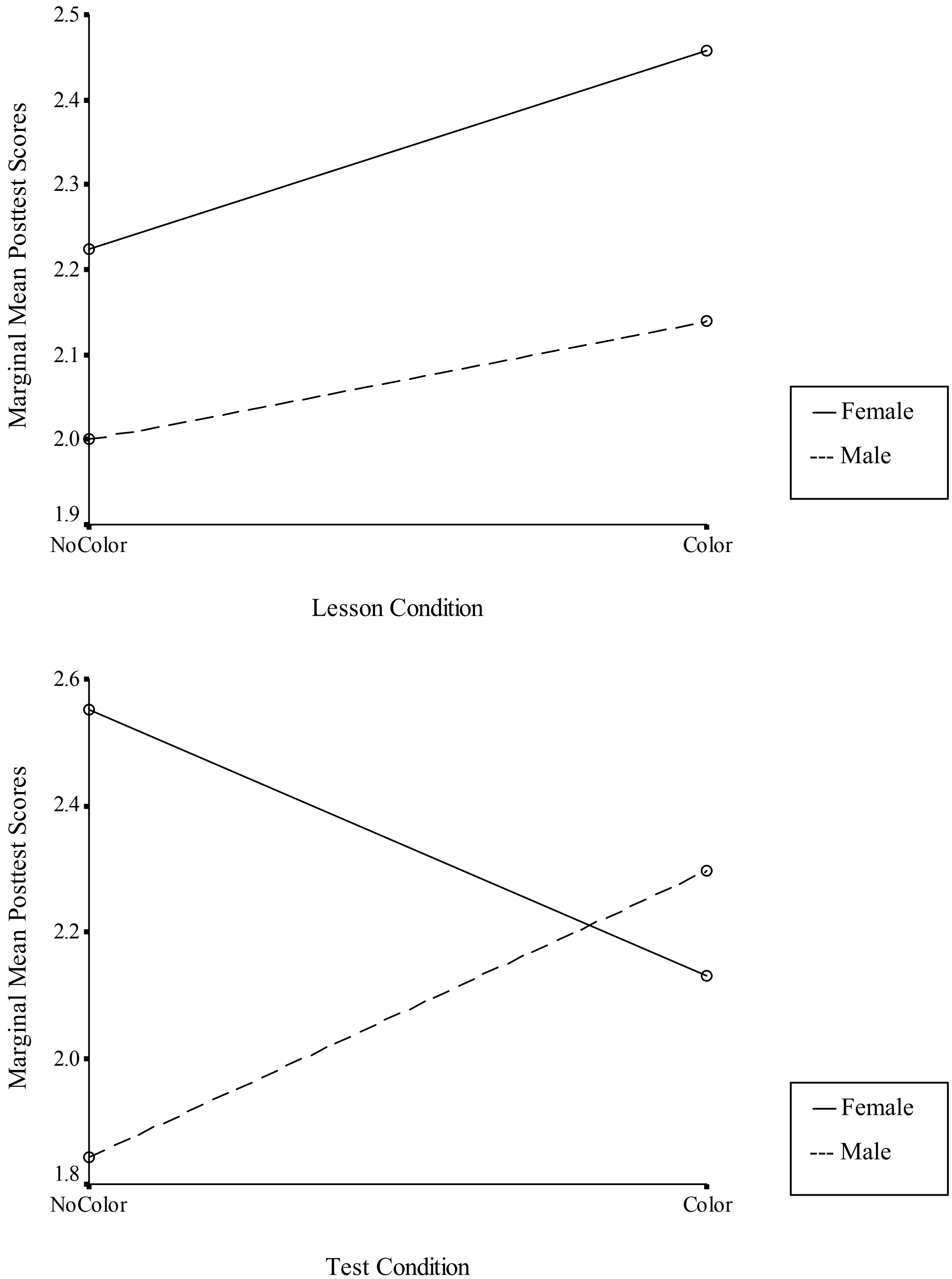

Figure 11. Plots for marginal means of posttest scores by gender for lesson and posttest condition (Note. Gender difference was not significant for lesson condition but was significant for posttest condition)

Table 11 also contains significant within-subjects interactions. Mauchly's test of sphericity (1940) indicated that the independence assumption had been violated for the within-subjects variable *item type*, $\chi^2(2) = 9.079$, $p = .011$, so the *df*-adjusted Greenhouse-Geisser statistics were reported in Table 11. Note that where significant differences were found in the non-*df*-adjusted, sphericity-assumed statistics, they were also present in the *df*-adjusted statistics.

Item type was significant, $F(1.889, 264.699) = 442.988$, $p < .001$, with participants recalling more FV items than FPT items and CK items. Item difficulty was significant, $F(1, 151) = 441.500$, $p < .001$, with participants recalling more easy items than hard ones. The interaction of item type and item difficulty was significant, $F(1.753, 264.699) = 110.699$, $p < .001$, with participants recalling more FVE items than any other. These significant interactions are of little interest in this investigation. More critically, the three-way interaction of item type, item difficulty, and posttest condition was also significant, $F(1.753, 264.699) = 3.199$, $p = .049$, with posttest color having an effect on FVE items, but the four-way interaction of item type, item difficulty, lesson condition, and posttest condition fell short of reaching significance, $F(1.753, 264.699) = 2.531$, $p = .089$.

To examine further the effects of item type and difficulty on recall of FVE items, a follow-up ANOVA was run, with posttest score as the dependent variable and treatment condition as the independent variable. The mean scores and effect sizes for the four treatment groups, NoC x NoC, NoC x C, C x NoC, and C x C, are shown in Table 12 and depicted in Figure 12. The follow-up ANOVA revealed that two treatment conditions had significant effects on recall of FVE items. Both effects were identical. Participants in both the *No Color x Color* group ($M = 5.90$, $SD = 2.158$, $n = 40$) and the *Color x No Color* group ($M = 5.90$, $SD = 2.479$, n

= 40) outscored the *No Color x No Color* control group (*M* = 4.81, *SD* = 1.991, *n* = 42). Post hoc Dunnett's t-tests compared each experimental treatment group against the control group and found the effect sizes for both the NoC x C and the C x NoC conditions (*ES* = .55) significant at the .05 level. While the color-matching group (C x C) had a positive effect (*ES* = .27), it was not statistically significant, contrary to encoding specificity principle predictions. When border colors were present in the lesson or the posttest, but not in both, color significantly enhanced recall of FVE items.

Table 12

Means and effect sizes for recall of FVE items

Lesson condition x Posttest condition	*M*	*SD*	*n*	Mean Difference	Effect Size
No Color x No Color	4.81	1.991	42	-	-
No Color x Color	5.90	2.158	40	1.09	0.55*
Color x No Color	5.90	2.479	40	1.09	0.55*
Color x Color	5.35	2.429	37	0.54	0.27
Total	5.48	2.292	159	-	-

Note: Maximum score for FVE portion of posttest = 9
* $p<.05$, ** $p<.01$

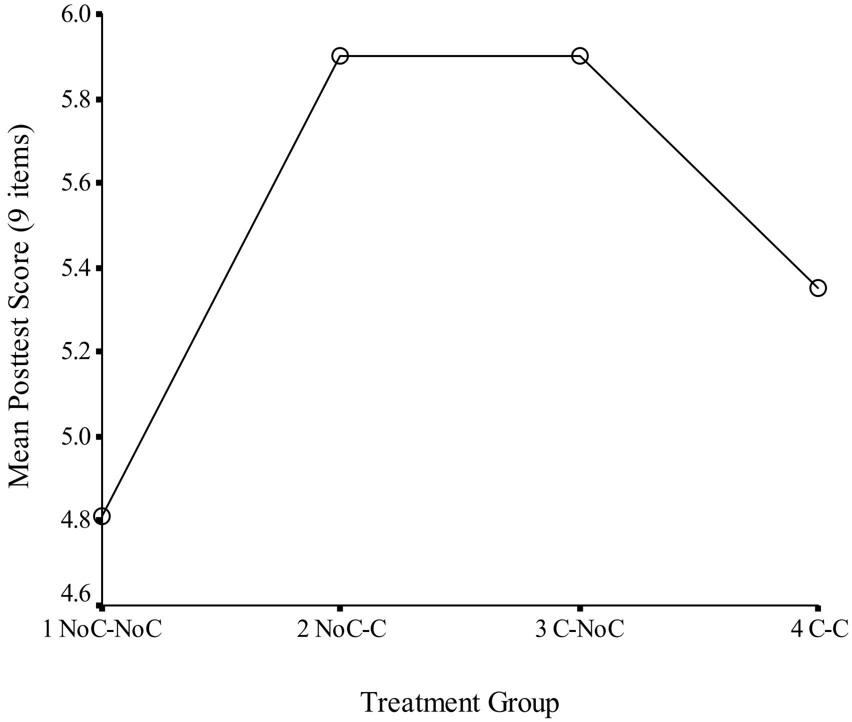

Figure 12. Plot for recall of FVE items by treatment condition with treatment groups 2 and 3 scoring significantly higher than the control group

Table 11 also contains a significant three-way interaction of item type, posttest condition, and gender, $F(1.889, 285.248) = 3.306$, $p = .041$. This is depicted with three plots in Figure 13. The presence of border colors in the posttest was moderately helpful for males across all measures and sub-measures but clearly unhelpful for females. To determine which effects were significant, a pair of follow-up ANOVAs were run, with posttest score as the dependent variable and posttest condition as the independent variable, measuring the effects of the presence of border colors in the posttest on all 10 criterion measures and sub-measures by gender. Overall, male participants recalled significantly more TBR content when border colors were present in the posttest (see Table 12), *ES* = .47, $F(1, 64) = 4.040$, $p = .049$, than when they were not present in the posttest.

Whether or not color was present in the lesson, color in the posttest significantly helped males in answering FV items, *ES* = .49, $F(1, 64) = 4.759$, $p = .033$, and more specifically, in answering FVE items, *ES* = .57, $F(1, 64) = 5.010$, $p = .029$. When border colors were present in the posttest ($M_C = 5.88$, $SD = 2.132$, $n = 33$), males recalled more FVE items than if there were no border colors in the posttest ($M_{NoC} = 4.58$, $SD = 2.292$, $n = 33$).

For females, the opposite occurred. Color effects were moderately negative across all measures and sub-measures. Female participants recalled 17.2% fewer items overall when border colors were present in the posttest. However, this over-arching negative effect failed to reach significance, *ES* = -.36, $F(1, 91) = 3.302$, $p = .072$ (see Table 13). Border colors in the posttest did significantly inhibit recall of FVH items, *ES* = -.46, $F(1, 91) = 5.638$, $p = .020$. When border colors were present in the posttest ($M_C = 2.18$, $SD = 1.618$, $n = 44$), females recalled 28.8% fewer FVH items than if there were no border colors in the posttest ($M_{NoC} = 3.06$, $SD = 1.919$, $n = 49$).

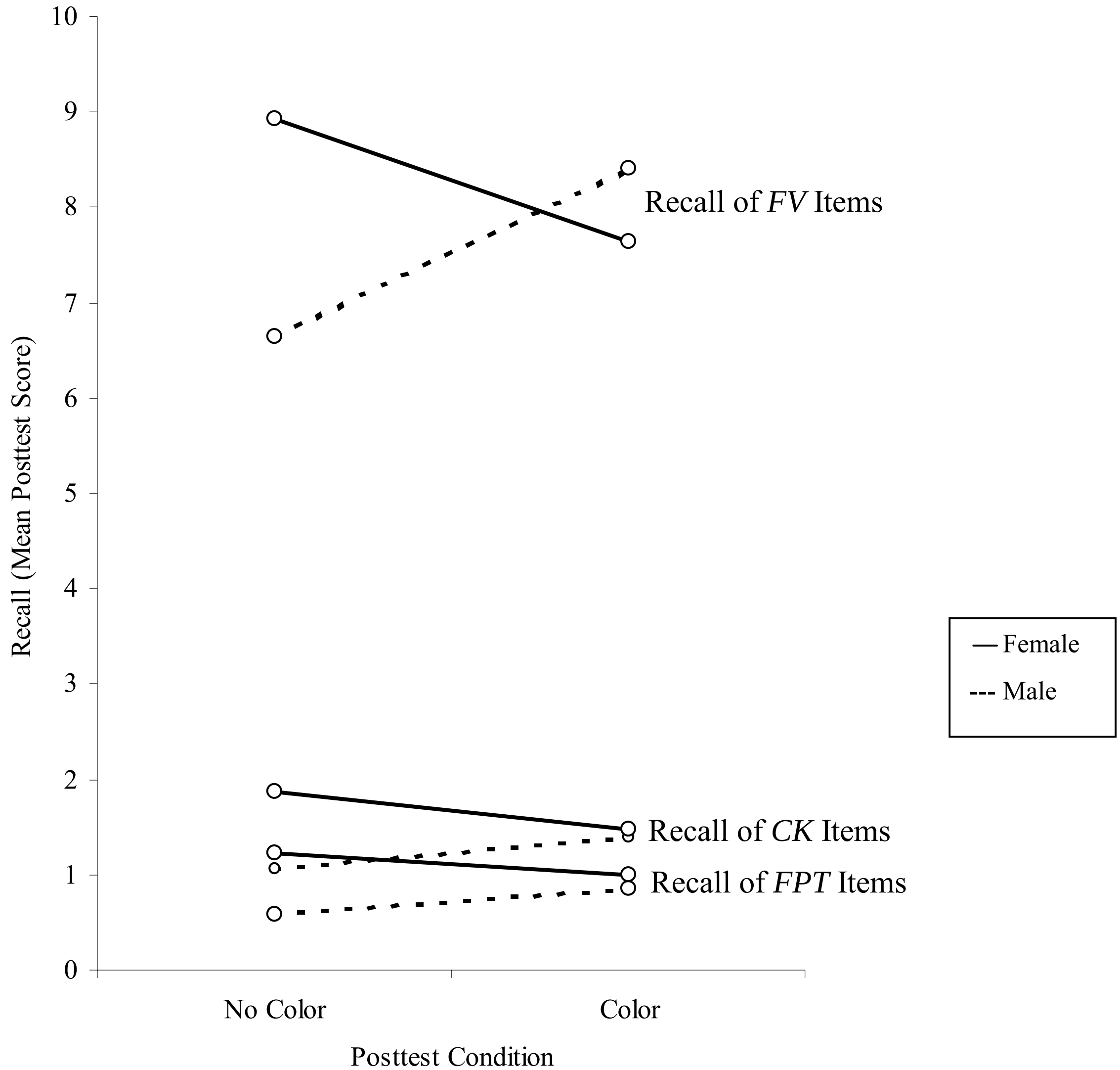

Figure 13. Plots of the effects of border colors in the posttest on recall by gender, including recall of factual-verbatim (FV), factual-paraphrased/transposed (FPT), and conceptual knowledge (CK) items

Table 13

2 x 10 ANOVA: Effects of border colors in the posttest for males

Source		*SS*	*df*	*MS*	*F*	*Sig.*
FVE	Between Groups	28.015	1	28.015	5.718*	0.020
	Within Groups	313.576	64	4.900		
	Total	341.591	65			
FVH	Between Groups	3.409	1	3.409	1.148	0.288
	Within Groups	190.121	64	2.971		
	Total	193.530	65			
FV	Between Groups	50.970	1	50.970	4.759*	0.033
	Within Groups	685.515	64	10.711		
	Total	736.485	65			
FPTE	Between Groups	4.909	1	4.909	1.543	0.219
	Within Groups	203.576	64	3.181		
	Total	208.485	65			
FPTH	Between Groups	1.227	1	1.227	1.086	0.301
	Within Groups	72.303	64	1.130		
	Total	73.530	65			
FPT	Between Groups	1.227	1	1.227	1.086	0.301
	Within Groups	72.303	64	1.130		
	Total	73.530	65			
CKE	Between Groups	0.545	1	0.545	0.751	0.389
	Within Groups	46.485	64	0.726		
	Total	47.030	65			
CKH	Between Groups	0.379	1	0.379	0.760	0.386
	Within Groups	31.879	64	0.498		
	Total	32.258	65			
CK	Between Groups	1.833	1	1.833	1.176	0.282
	Within Groups	99.758	64	1.559		
	Total	101.591	65			
Total	Between Groups	139.636	1	139.636	4.040*	0.049
	Within Groups	2212.303	64	34.567		
	Total	2351.939	65			

* $p < .05$, ** $p < .01$

Table 14

2 x 10 ANOVA: Effects of border colors in the posttest for females

Source		*SS*	*df*	*MS*	*F*	*Sig.*
FVE	Between Groups	3.758	1	3.758	0.717	0.399
	Within Groups	476.909	91	5.241		
	Total	480.667	92			
FVH	Between Groups	17.929	1	17.929	5.638*	0.020
	Within Groups	289.362	91	3.180		
	Total	307.290	92			
FV	Between Groups	38.102	1	38.102	2.995	0.087
	Within Groups	1157.855	91	12.724		
	Total	1195.957	92			
FPTE	Between Groups	13.251	1	13.251	3.409	0.068
	Within Groups	353.738	91	3.887		
	Total	366.989	92			
FPTH	Between Groups	1.168	1	1.168	0.582	0.447
	Within Groups	182.531	91	2.006		
	Total	183.699	92			
FPT	Between Groups	1.168	1	1.168	0.582	0.447
	Within Groups	182.531	91	2.006		
	Total	183.699	92			
CKE	Between Groups	2.081	1	2.081	2.371	0.127
	Within Groups	79.876	91	0.878		
	Total	81.957	92			
CKH	Between Groups	0.149	1	0.149	0.197	0.658
	Within Groups	68.840	91	0.756		
	Total	68.989	92			
CK	Between Groups	3.345	1	3.345	1.646	0.203
	Within Groups	184.977	91	2.033		
	Total	188.323	92			
Total	Between Groups	161.868	1	161.868	3.302	0.072
	Within Groups	4460.261	91	49.014		
	Total	4622.129	92			

* $p < .05$, ** $p < .01$

Hypothesis 7: Mismatching color context effects

This hypothesis predicted that the mismatching color condition, in which there were border colors in the lesson and intentionally mismatched border color hues in the posttest, would inhibit recall. Participants in the color-mismatching condition (Color x MisColor) should have scored significantly lower than the color-matching condition (Color x Color). Given the gender differences found in the previous analysis, recall on all 44 posttest items was compared by gender for the color matching and color mismatching groups. This 2 x 2 ANOVA (treatment group x gender), with total posttest score as the dependent measure, yielded no significant treatment or gender differences (see Table 15). In absolute terms, mismatching the border colors had no impact at all on females ($M_{\text{C-C}} = 13.57$, $SD = 6.860$, $n = 21$; $M_{\text{C-MisC}} = 13.22$, $SD = 6.612$, $n = 23$; $ES = -.05$). For males, there was a small negative difference ($M_{\text{C-C}} = 13.24$, $SD = 5.901$, $n = 17$; $M_{\text{C-MisC}} = 11.76$, $SD = 5.437$, $n = 17$; $ES = -.25$), suggesting that males were slightly more sensitive to the matching and mismatching pattern of the hues than females. However, since the negative effects of mismatching the border colors were not significant, the null cannot be rejected.

Table 15

2 x 2 ANOVA: Total posttest scores for matching and mismatching groups by gender

Source	*SS*	*df*	*MS*	*F*	*Sig.*
Treatment Group (TG)	15.949	1	15.949	.402	.528
Gender (G)	15.329	1	15.329	.387	.536
TG * G	5.972	1	5.972	.151	.699
Error	2933.174	74	39.637		
Total	16150.000	78			

Note. The total posttest score measure involved all 44 posttest items
TG = Treatment Group: C x C and C x MisC

Exploratory Analysis

Topic-color matching test

After completing all of the lessons and posttests, participants who had seen border colors in the lesson and/or the posttest were given a topic-color-matching (TCM) test, in which they were asked to identify the color that corresponded with a particular topic. The eight hues used as border colors in the course were provided as options. Each hue corresponded with one of the eight main menu topics. Participants in the control group were not given the TCM test since they had no way of knowing which color went with which topic. Pure guessing would have yielded a score of 1 out 8, or 12.5% correct. Since a gender interaction with color was detected in earlier analysis, the TCM scores were analyzed by gender.

A 2 x 2 ANOVA (treatment condition x gender), with TCM scores as the dependent variable, was run but no significant treatment interactions were found. There was no significant three-way interaction for gender, treatment condition, and TCM scores, $F(3, 146) = 1.552$, $p = .204$. With one exception, no group was able to score much higher than 12.50%, the level of chance guessing (see Table 16 and Figure 14). Specifically, female participants in the color-matching group (Color x Color) were, on average, able to match 21.38% of the topics and colors ($M_{F,C\text{-}C} = 1.71$, SD = 1.454, n = 21). In comparison to the color-mismatching group ($M_{F,C\text{-}M} = .78$, SD = 1.085, n = 23), the difference was significant ($ES = .86$, $p = .043$). Based on these results, males did not explicitly remember the topic-color associations, while females remembered them only when cued with border colors in both the lesson and the posttest. Even so, remembering more of the color information did not necessarily help females in the color-matching treatment group recall more of the TBR information. Overall, it can be concluded that participants did not explicitly (or consciously) recognize the topic-color associations much beyond the level of chance guessing. Correlation between TCM test scores and posttest scores was very low, although positive (Pearson's r = .187 overall, .143 for females, and .036 for males), suggesting that explicitly remembering topic-color (content-context) information did not necessarily contribute to improved recall of the TBR content.

Table 16

Means scores for the Topic-Color Matching (TCM) Test

	Male		Female		Total	
Lesson x Posttest	*n*	*M* *SD*	*n*	*M* *SD*	*n*	*M* *SD*
No Color x Color	17	1.29 1.263	24	1.00 1.022	41	1.12 1.122
Color x No Color	15	.67 .816	27	1.04 1.055	42	.90 .983
Color x Color	17	1.00 .707	21	1.71 1.454	38	1.39 1.220
Color x MisColor	17	1.06 .899	23	.78 1.085	40	.90 1.008
Total	66	1.02 .953	95	1.12 1.184	161	1.07 1.093

Max score = 8
Note: The control group (No Color x No Color) was not given the TCM Test to complete

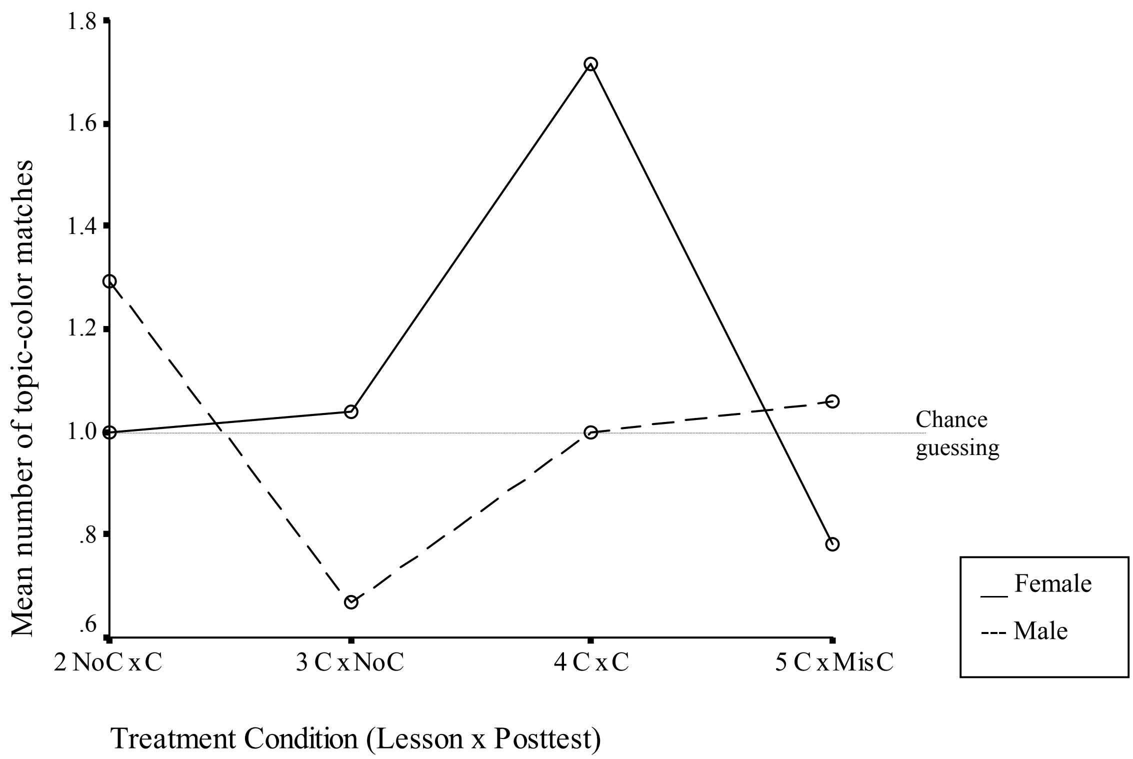

Figure 14. TCM scores by treatment group for male and female participants (NoC x NoC control group was not given the TCM test).

Lesson duration

To determine if reading speed varied by lesson condition, posttest condition, and/or gender, a 2 x 2 x 2 MANOVA (lesson condition, posttest condition, and gender) was run, with lesson duration (in seconds) as the dependent measure. There were eight lessons so there were eight lesson duration measures. Only one significant difference was found, the two-way interaction of gender and lesson condition was significant for lesson 6, $F(1, 151) = 4.291$, $p = .040$ (see Table 17). Males took 26 seconds longer on average to read lesson 6 when color was present in the lesson, while females took 11 seconds less to read lesson 6 when color was present in the lesson. Although statistically significant, in practical terms these differences are small and appear in only one of the eight lessons.

Table 17

8 x 2 x 2 x 2 MANOVA: Lesson duration (reading speed) by lesson condition by posttest condition x gender

Source	*SS*	*df*	*MS*	*F*	*P*
Lesson Condition (L)					
Duration - Lesson 1	570.829	1	570.829	.294	.588
Duration - Lesson 2	5855.122	1	5855.122	2.325	.129
Duration - Lesson 3	22.030	1	22.030	.011	.917
Duration - Lesson 4	9897.359	1	9897.359	1.754	.187
Duration - Lesson 5	5466.508	1	5466.508	1.778	.184
Duration - Lesson 6	9858.714	1	9858.714	3.224	.075
Duration - Lesson 7	3180.554	1	3180.554	.943	.333
Duration - Lesson 8	14363.323	1	14363.323	3.253	.073
Posttest Condition (P)					
Duration - Lesson 1	4175.949	1	4175.949	2.153	.144
Duration - Lesson 2	2643.070	1	2643.070	1.049	.307
Duration - Lesson 3	234.266	1	234.266	.115	.735
Duration - Lesson 4	113.390	1	113.390	.020	.887
Duration - Lesson 5	3451.517	1	3451.517	1.123	.291
Duration - Lesson 6	2103.760	1	2103.760	.688	.408
Duration - Lesson 7	482.739	1	482.739	.143	.706
Duration - Lesson 8	1968.495	1	1968.495	.446	.505
Gender (G)					
Duration - Lesson 1	972.150	1	972.150	.501	.480
Duration - Lesson 2	79.023	1	79.023	.031	.860
Duration - Lesson 3	29.589	1	29.589	.014	.904
Duration - Lesson 4	2177.184	1	2177.184	.386	.535

Duration - Lesson 5	114.177	1	114.177	.037	.847
Duration - Lesson 6	951.168	1	951.168	.311	.578
Duration - Lesson 7	1301.843	1	1301.843	.386	.535
Duration - Lesson 8	848.402	1	848.402	.192	.662
L x P					
Duration - Lesson 1	4402.303	1	4402.303	2.270	.134
Duration - Lesson 2	3124.556	1	3124.556	1.241	.267
Duration - Lesson 3	1327.918	1	1327.918	.650	.422
Duration - Lesson 4	4278.067	1	4278.067	.758	.385
Duration - Lesson 5	12.304	1	12.304	.004	.950
Duration - Lesson 6	1.289	1	1.289	.000	.984
Duration - Lesson 7	5057.990	1	5057.990	1.500	.223
Duration - Lesson 8	6071.911	1	6071.911	1.375	.243
L x G					
Duration - Lesson 1	6087.114	1	6087.114	3.138	.078
Duration - Lesson 2	2874.832	1	2874.832	1.141	.287
Duration - Lesson 3	3672.427	1	3672.427	1.796	.182
Duration - Lesson 4	1879.568	1	1879.568	.333	.565
Duration - Lesson 5	4733.116	1	4733.116	1.540	.217
Duration - Lesson 6	4762.276	1	4762.276	1.557	.214
Duration - Lesson 7	.129	1	.129	.000	.995
Duration - Lesson 8	6537.513	1	6537.513	1.481	.226
P x G					
Duration - Lesson 1	5415.094	1	5415.094	2.792	.097
Duration - Lesson 2	3982.753	1	3982.753	1.581	.211
Duration - Lesson 3	6149.151	1	6149.151	3.008	.085
Duration - Lesson 4	7481.962	1	7481.962	1.326	.251
Duration - Lesson 5	10275.301	1	10275.301	3.343	.069
Duration - Lesson 6	13122.377	1	13122.377	4.291*	.040
Duration - Lesson 7	3212.384	1	3212.384	.953	.331
Duration - Lesson 8	942.003	1	942.003	.213	.645
L x P x G					

Duration - Lesson 1	16.461	1	16.461	.008	.927
Duration - Lesson 2	105.124	1	105.124	.042	.838
Duration - Lesson 3	1052.600	1	1052.600	.515	.474
Duration - Lesson 4	21.199	1	21.199	.004	.951
Duration - Lesson 5	9859.449	1	9859.449	3.207	.075
Duration - Lesson 6	375.622	1	375.622	.123	.726
Duration - Lesson 7	158.971	1	158.971	.047	.828
Duration - Lesson 8	733.940	1	733.940	.166	.684
Error					
Duration - Lesson 1	292869.432	151	1939.533		
Duration - Lesson 2	380327.281	151	2518.724		
Duration - Lesson 3	308703.564	151	2044.394		
Duration - Lesson 4	851992.465	151	5642.334		
Duration - Lesson 5	464159.006	151	3073.901		
Duration - Lesson 6	461793.710	151	3058.236		
Duration - Lesson 7	509075.300	151	3371.360		
Duration - Lesson 8	666632.443	151	4414.784		
Total					
Duration - Lesson 1	2521603.293	159			
Duration - Lesson 2	2855999.096	159			
Duration - Lesson 3	2248982.638	159			
Duration - Lesson 4	4504430.323	159			
Duration - Lesson 5	3845655.253	159			
Duration - Lesson 6	3283639.399	159			
Duration - Lesson 7	2922115.949	159			
Duration - Lesson 8	3370254.021	159			

* $p < .05$, ** $p < .01$

Summary of Results

In the data analysis, significant gender differences were found for color effects on posttest scores. For males, having border colors in the lesson enhanced recall, but not significantly (*ES* = .19). However, the presence of border colors in the posttest enhanced recall for males significantly (*ES* = .47). The influence of border colors on recall was most apparent in the FV (factual verbatim) items (*ES* = .49), and more specifically, in the FVE (factual verbatim easy) items (*ES* = .57).

For females, border colors in the lesson also enhanced recall, but not significantly (*ES* = .22). Conversely, the presence of border colors in the posttest had a negative effect (*ES* = -.36), although this also did not reach significance. However, the negative effects for border color in the posttest did reach significance for FVH (factual verbatim hard) items (*ES* = -.46).

The main treatment effects were not significant. The *Color x Color* group (*ES* = .14) recalled more than the no color x color group (ES = .06) but failed to outperform the *Color x No Color* group (*ES* = .30). The only significant treatment effect found was for males in the *No Color x Color* group (*ES* = 1.22). Intentionally mismatching the border color hues in the posttest with those shown in the lesson yielded non-significant negative effects (ES_{female} = -.05 and ES_{male} = -.25), relative to the color-matching (color x color) group.

In the exploratory analysis, it was found that females in the color-matching group were able to identify significantly more topic-color interactions, while all other groups scored close to or below the level of chance guessing. With regard to reading speed, there were no consistently significant differences favoring either gender, either lesson condition, or either posttest condition. The presence of border colors had no impact on reading speed. The results of the data analysis will be discussed at length in the next chapter in relation to the hypotheses and the relevant literature.

Chapter 5

DISCUSSION

The main purpose of this investigation was to examine the effects of contextual border colors presented on-screen in a computer-based instruction module. In this section, the findings for each research question are discussed. This is followed by a general discussion of the findings, their implications, the limitations of these results, and suggestions for future research.

Findings

Research Question 1

Does the presence of border colors enhance recall task performance? Although the general trend of the results indicates that participants who completed a version of the course containing border colors (M_{Color} = 12.15, *SD* = 5.688, *n* = 117), particularly those with border colors in the lesson, generally scored higher than the control group ($M_{NoColor}$ = 11.07, *SD* = 6.190, *n* = 42), neither of the between-subjects differences in the global MANOVA were significant. Therefore, the null hypothesis cannot be rejected. Border colors in the lesson had a non-significant effect on the recall posttest results. Border colors in the posttest had no effect overall (*ES* = -.05). However, as will be discussed, offsetting effects for male and female participants obscured the effects of border colors in the posttest.

Research Question 2

Does *matching* the border color displayed in the posttest with that displayed in the lesson optimize posttest achievement? The *Color x Color* group (M_{CxC} = 13.46, *SD* = 6.449, *n* = 37) did recall more information than the control group ($M_{NoCxNoC}$ = 12.48, *SD* = 7.134, *n* = 42). However, this difference did not reach significance. In addition, the *Color x Color* group failed to outperform the *Color x No Color* group (M_{CxNoC} = 14.65, *SD* = 7.294, *n* = 40). Therefore, we cannot reject the null hypothesis and must conclude that matching the border color in the posttest with that shown in the lesson did not optimize the context effects of color on recall. These results run counter to the encoding specificity principle (Tulving & Thomson, 1973), which predicted that matching the posttest conditions with the lesson conditions would result in significant benefits from context effects. Also, the fact that the *Color x No Color* group recalled more than all other groups indicates that encoded border colors had an effect on recall even though they were not present in the posttest. In reference to environmental context cues, Smith (1988) posits that individuals can imagine a previously encoded memory cue and use that internally generated image as a retrieval cue, even when the cue is not physically present (p. 23). However, this internal manipulation of context effect (*ES* = .30) failed to reach significance, therefore no conclusions can be drawn from it.

Research Questions 3, 4, and 5

Is there an interaction between the type of factual recall task assessed (involving verbatim and paraphrased/transposed test items) in the posttest and the context effects of border color? Does recall of factual knowledge vary from that of higher-level conceptual knowledge in terms of the context effects of color? Is there an interaction between posttest item difficulty (easy and difficult) and the context effects of border color?

Yes, both item type and item difficulty yielded significant interactions. Participants, not surprisingly, scored significantly higher on easy items than on difficult ones. More importantly, participants recalled a greater number of *factual-verbatim* items (M_{FV} = 7.98 or 44%, *SD* = 3.519, *n* = 159) than *factual-paraphrased/transposed* items (M_{FPT} = .95 or 5%, *SD* = 1.292, *n* =

159) and *conceptual knowledge* items (M_{CK} = 1.49 or 19%, *SD* = 1.373, *n* = 159). These item type differences are examined in light of fuzzy-trace theory (Brainerd & Reyna, 1990) in the general discussion.

Although the interaction of the within-subjects variables *item type* and *item difficulty* were significant, there were no significant two-way interactions between them and lesson or posttest condition. However, there was a significant three-way interaction among item type, item difficulty, and posttest condition. Follow-up analysis revealed that when border colors were present in the lesson or in the posttest, recall of FVE items was significantly higher (*ES* = .55) for the treatment groups than for the control group. Recall of FVE items was not significantly higher when border colors were presented in both the lesson and the posttest (*ES* = .27).

In summary, item type and item difficulty did interact with the context effects of border colors. Thus, we can reject the null hypotheses and accept the alternate hypotheses: the type of recall task (i.e., FV versus FPT versus CK item types) and the difficulty of the TBR content *do* influence the context effects of border colors in computer-based instruction. Recall of *factual-verbatim-easy* information (FVE) is enhanced by the presence of border colors in either the lesson or the posttest (but not necessarily both). This finding is consistent with fuzzy-trace theory (Brainerd & Reyna, 1990) but is not consistent with *outshining hypothesis* (Smith et al., 1978). This will be examined further in the general discussion.

Research Question 6

Is there an interaction between gender and the context effects of color? The MANOVA revealed a significant difference between male and female posttest performance. Females consistently outperformed males in recall of FV, FPT, and CK items (*ES* = .72). This is consistent with several studies indicating that females have a significant advantage over males in answering constructed response test items (Bolger & Kellaghan, 1990; Breland et al., 1994; Bridgeman & Lewis, 1994; De Mars, 1998, 2000; Mazzeo et al., 1993; Murphy, 1982; Weaver & Raptis, 2001).

The MANOVA also yielded a significant three-way interaction for gender, item difficulty, and posttest condition. Border colors in the *posttest* had notably different effects on males and females. When border colors were present in the posttest, males recalled significantly more information (*ES* = .47), while females did not (ES = -.36). In fact, females recalled significantly less FVH information (ES = -.46). Since the gender-color interaction was significant, the null hypothesis may be rejected. Border colors *do* have differential context effects for male and females when presented at retrieval. This finding will be examined more closely in the general discussion.

Research Question 7

Does *mismatching* the border color hues displayed in the posttest with those displayed in the lesson undermine posttest achievement? No, although mismatching the border colors did inhibit total recall slightly for males, the difference was not significant. Nevertheless, the trend that mismatching appears to have some impact on recall for males but none for females is consistent with the other gender-related differences found. Males benefited significantly from border colors in the posttest but were also somewhat negatively affected by mismatching color cues in the posttest. For females, recall was suppressed somewhat by matching border colors in the posttest, but this negative effect was no better or worse when the border colors contained mismatching hues.

Exploratory Analysis – Topic-Color Association

Overall, neither male nor female participants explicitly remembered the topic-color associations, although females generally remembered the topic-color information more than males. Females in the *Color x Color* group recognized topic-color associations significantly better than those in the (*Color x MisColor*) outside control group. However, even the female color-matching group was not able to recognize more than two sets of topic-color associations, out of eight. In the Patel, Blades, and Andrade study (1999), participants remembered 6 out of 8 object-color associations, on average. However, note that the object-color associations reflected memory of *concrete* objects (e.g., a green comb), whereas in the present study, participants were tested for *abstract* associations between topics and colors (e.g., green expectancy theory).

The low TCM test scores are not surprising, considering that participants were never informed that either the color information or the topic-color associations would be important to remember. As such, participants were not likely to have converted the topic-color associations into verbal propositions (i.e., "This color corresponds with that topic"), a step that would presumably be necessary in order to answer TCM Test questions correctly. This suggests that while color may have been encoded implicitly, as Hasher and Zacks' (1979) automaticity hypothesis posits, the topic-color *associations* may not have been encoded. Another possible explanation for the low TCM scores is that perhaps associations between concepts and colors have weaker activation strengths than associations between physical objects and colors. A simpler explanation is that the instrument used in the present investigation was simply not sensitive enough to capture color memory, since it measured topic-color associations and not color recognition directly. Clariana (2003) found that constructed response tasks inhibited memory of the color scheme used, suggesting that in the current study, the fill-in-the-blank posttest (a constructed response task) suppressed color memory for the subsequent topic-color matching test.

Murnane and his colleagues (Murnane, Phelps, & Malmberg, 1999) suggest that encoding topic-color associations – what they call ensemble information – requires active, explicit encoding, not the passive, implicit encoding involved in the current study. The lack of explicit topic-color ensemble knowledge may partially explain why the *Color x Color* group did not recall significantly more TBR information than the other groups, contrary to the encoding specificity principle (Tulving & Thomson, 1973).

Exploratory Analysis – Reading Speed

The results generally indicate that the presence of color in the lesson had negligible impact on reading speed and no impact at all when color was present in the posttest. In other words, it does not appear that participants took extra time to encode associations between TBR content and the border colors. The presence of color was not a significant distraction in terms of reading speed. Consistent with Hasher and Zacks' (1979) automaticity hypothesis, to the extent that it impacts learning, color appears to be encoded implicitly without taxing the learner's working memory or, consequently, the learner's reading speed.

General Discussion

Context effects of color

If border colors had strong context effects on recall, the encoding specificity principle (Tulving & Thomson, 1973) would predict that participants in the color-matching group would significantly outscore the control group and all other mismatching groups. If the context dependency of recall were strong and color was a powerful context variable, one would also predict that mismatching the color hues shown in the posttest with those shown in the lesson would inhibit recall. Although the trend of the data suggests that the presence of border colors in the lesson had an effect on recall (*ES* = .22), neither hypothesis could be affirmed. Border color, being external to the content rather than intermingled with the content, may simply not be a strong enough context variable to generate significant effects on this form of recall posttest (i.e., a difficult fill-in-the-blank test). It is also possible that color in general is too pervasive and too fragile a contextual variable to elicit strong context effects (i.e., semantic cues, practice effects, and other variables outshine color cues), particularly when the learner is not aware that color is important. Verbal context cues have the advantage that they can provide a unique context for each TBR word and they influence the meaning of associated TBR information (Smith, 1986), whereas non-verbal context cues usually must be associated with multiple TBR words and do not necessarily influence meaning. In the present investigation, a single color hue was associated with a *group* of concepts related to a topic and there was no obvious association between the color hues and the topics. Thus, the explicit influence of color context was minimal.

In any case, the results of the present study are *not* consistent with the encoding specificity principle (Tulving & Thomson, 1973), and lend support to some of its critics (Epstein & Dupree, 1977; Epstein et al., 1979; Martin, 1975; Pellegrino & Salzberg, 1975; Santa &

Lamwers, 1974, 1975). Cuing retrieval with the same border color that was present during encoding did not significantly enhance recall. The present findings can be interpreted as being more consistent with a *generation-recognition* model of memory (Bahrick, 1970). Within this model, the learner's ability to generate possible answers to the fill-in-the-blank questions and to recognize the correct answer is independent of contextual variables. The presence of context cues, e.g., border colors, should have no significant effect on recall task performance (Epstein & Dupree, 1977). The fact that none of the treatment groups performed significantly better than the control group suggests that color context had no impact on recall, thus supporting a *generation-recognition* interpretation. However, *some* significant color effects *were* found once the data was analyzed by item type, item difficulty, and gender. The *generation-recognition* model does not account for such differences. Thus, the results do not lend support to either the encoding specificity principle or the generation-recognition model.

Despite the lack of statistically significant differences, the general trend of the results indicates that color in the lesson enhanced recall for both males and females and that color in the posttest helped males recall more TBR content while inhibiting recall for females. The small effect sizes may be due to the limitations of the measurement instruments, rather than being reflective of the actual impact of color on memory.

Item format interactions

Verbatim traces appear to be stronger than gist traces when tested using fill-in-the-blank items shortly after reading a passage. This is consistent with fuzzy-trace theory, which predicts that verbatim memory traces are stronger than gist traces immediately following encoding (Brainerd & Reyna, 1990). Also, the finding that recall of gist items was about the same as recall of conceptual knowledge reaffirms Clariana and Koul's (2002) argument that paraphrasing and transposing constructed response items demands of the learner a conceptual understanding of the underlying associations being tested that goes beyond simple recall of facts. Thus, FPT item types may border or even overlap with conceptual knowledge test items, at least in terms of activation strength.

It appears that with *easy* factual information (FVE), recall is influenced by the presence of border colors in the lesson or the posttest (but not both). On the other hand, context cues appear to have little impact on verbatim traces for *difficult* information (FVH) or gist traces of any sort (FPT and CK items). These results appear to run counter to those in a study by Clariana (2002), in which color improved recall of difficult items by 20% but did not improve recall of easy items. This effect (*ES* = .35) was not statistically significant. Nevertheless, if one considers that the "difficult" items in Clariana's study had probabilities of .53 (color) and .44 (no color), and the "easy" items in the current study had probabilities of .66 (color) and .53 (no color), then one might conclude that test items within this difficulty range (roughly, 44% to 53%) are susceptible to the effects of color. Just as Goldilocks preferred her porridge neither too hot nor not too cold, perhaps color context is most influential when fill-in-the-blank items are neither too easy nor not too difficult.

The *outshining hypothesis* (Smith et al., 1978) predicted that color context information would have its greatest impact on the most difficult items and the weakest effects on the easiest items. In truth, with fill-in-the-blank posttest items, there were very few truly easy items. One might argue that the present study contained difficult and *more* difficult items rather than easy and difficult. Accepting that argument, the results of the present study suggest that color context will help learners the most for items that are moderately difficult, with activation strengths of roughly .5, and may be less effective with items that are too difficult (activation strengths well below .5) or too easy (activation strengths well above .5). Further research should explore activation strength further in relation to color context effects to determine a more precise range within which context influences recall.

The results are consistent with *fuzzy-trace theory* (Brainerd & Reyna, 1990), insofar as color context enhanced recall of verbatim traces while having less of an impact on gist traces. The results might have been different with a delayed test of recall, which would have allowed time for verbatim traces to fade. Similarly, if the lesson had involved greater opportunity for elaboration, e.g., through embedded practice exercises, the probability of forming and strengthening gist memory traces should have improved. Instead, by having the participants simply read the passages and immediately complete the fill-in-the-blank test, there was little opportunity for gist memory traces to form. Thus, the finding that color influences verbatim memory more than it does gist memory should be accepted with caution.

Attention-Arousal and Cognitive Load

This study found significant gender-color interactions. What could explain the significant improvement for males and the non-significant negative effects for females? Perhaps the answer lies in how color is perceived by the individual.

Berlyne (1960) suggests that the collative properties, i.e., the novelty, complexity, and incongruity of the visual stimuli, can elicit arousal, and that arousal can govern how interesting learners find the material and therefore how much attention they allocate to it (Berlyne, Craw, Salapatek, & Lewis, 1963). Kahneman (1973) proposes that there is a positive correlation between attention and effort. In the present investigation, the novelty and incongruity of the border colors in the posttest, with their changing hues, may have aroused the attention of males, causing them to put more effort into the difficult task of recalling the key terms. If it is true that the color elicited *attention-arousal effects* in males, why were females negatively influenced, when influenced at all, by the presence of border colors in the posttest?

An alternate explanation is rooted in cognitive load and implicit memory research. Some studies suggest that color information is encoded automatically without incurring taxing cognitive load (Backman et al., 1993; Hasher & Zacks, 1979; Hatwell, 1995; Ling & Blades, 1996). If male subjects were *not* aware of the posttest border colors changing hue in the periphery of the screen, they would *not* have taxed working memory trying to process the color information. The color information would have been encoded automatically and yielded *attention-arousal effects* without taxing memory. Conversely, if females were aware of the presence of color, they might have actively encoded the color information. The color information may then have competed with the true TBR content at retrieval, creating cognitive load issues that overwhelmed any *attention-arousal* benefits they might have realized. Supporting this argument is the fact that females performed better than males on the topic-color matching test, suggesting that they actively tried to remember the topic-color associations, while males encoded them implicitly. Additional research will be required to determine if presenting changing border colors in the posttest indeed caused *attention-arousal effects* for males and *cognitive overload effects* for females, and if so, why.

Affordances of Color

The present findings have been described in cognitive terms, yet they can also be seen through an ecological lens. Heavily influenced by Gestalt theorists, such as Wertheimer (1959), Gibson's (1966) information pick-up theory posits that perception relies on information available in the stimulus array, rather than on purely internal mental processes, and that the observer directly acquires cues from physical objects. According to Gibson (1977), physical objects in the world have affordances. That is, they afford the individual with ways of interacting with them. For example, a "sit-on-able surface" (e.g., a chair) affords sitting, a "stand-on-able surface of support" (e.g., a floor) affords rest and standing (Gibson, 1982, p. 1). Note that this affordance is not merely a physical property of the chair or table nor only a cognitive manifestation, rather it is the interaction of those physical properties with the perceptions of the observer (Gibson, 1979). When the observer takes action and exercises an affordance (e.g., sits down in the chair), that affordance becomes an *effectivity* (Ryder & Wilson, 1996). It is this ecological, observer-centered, *observer-interacting-in-a-physical-world* approach that distinguishes Gibson's work with perception.

Norman (1988) popularized the notion of *affordances*, applying it to user-centered design. Norman argued that good design (1) provides the user with a conceptual model for how the object works and (2) allows clear visibility so that the user can perceive the affordances, constraints, and conventions involved. Norman (1999) later amended this, distinguishing between *real* affordances, those reflecting the physical reality of the world, and *perceived* affordances, those reflecting the observer's subjective perceptions of that reality. *Perceived* affordances, are particularly relevant to computer-based instruction, where many affordances and effectivities are *virtual*, not physical. For example, when a learner moves a mouse across the mouse pad, a pointer moves across the screen. The mouse physically moves, but the pointer's movement is nothing more than a visual illusion. Yet, to the learner, the pointer's movements are real, and the virtual constraints and affordances of the pointer are perceived to be as real as those of the mouse. The same distinctions can be drawn between typing on a keyboard and characters appearing on-screen and between clicking the mouse while the mouse pointer appears to be situated over an area designated as a button or hot spot.

What about the color then? What are its real and perceived affordances? Color is a product of the observer's intake and processing of light waves being absorbed and transmitted by physical objects. Since different objects absorb and transmit light differently, color hue is a distinctive property of physical objects. One could argue then that color affords the observer a way of distinguishing objects from backgrounds (Gibson, 1966). Berlyne and his colleagues (Berlyne, Ogilive, & Parham, 1968) suggest that color can also afford arousal to the observer, as a visual stimulus.

However, color is used in human-computer interface (HCI) as a means of color-coding semantic information, as with symbology (Kludt & Dyre, 2001; Post, Geiselman, & Goodyear, 1999). In educational research, color is often used to code different categories of conceptual information (Dwyer & Moore, 1994; Dwyer & Moore, 2001; Lamberski & Dwyer, 1983; Moore & Dwyer, 1997). In both cases, the color-coding is most often arbitrary, intended to help users/learners differentiate and group color-coded content based on an external concept mapping of what colors correspond with what concepts. Color-coding then could be said to offer the perceived affordances of grouping related concepts and providing structure to content. Perceived affordances, however, must be explicitly understood by learners in order to be affordances, whereas real affordances do not (Norman, 1999). If a fighter pilot does not know that some colors connote fire control information while others group navigational information, the color-coding is not perceived, and is therefore not a perceived affordance.

In the current investigation, participants were not made aware of the color-coding convention employed. Therefore, color did not have the perceived affordances of grouping concepts and providing content structure. Only the real affordances were available to the participants. Consistent with Gibson (1966), color acted as an invariant in the ambient visual array, adding distinctiveness to different portions of the course, with differing outcomes for males and females. This suggests that the presence of border colors at retrieval offers different attention-arousal affordances for males and females.

Implications of the Study

Taken at face value, the results of the present study suggest that presenting learners with border colors at retrieval may help narrow the gap between males and females on constructed response tasks, by enhancing recall for males while suppressing recall for females. This is obviously not a very practical instructional design strategy. A more pragmatic approach would be to optimize recall for *both* males and females by altering the use of color context cues at retrieval by gender. For males, having border colors at retrieval that change in hue for each test item based on a color-coding scheme may improve recall of TBR information, possibly by arousing attention and increasing effort. For females, it is prudent to avoid changing border color hues. However, these implications should be viewed with skepticism until these findings can be replicated with other studies and the causes can be more fully understood.

The gender differences found in the current study imply that men and women may differ in the way they perceive and process contextual stimuli. While both genders may encode contextual information automatically, they may vary in terms of how they use context cues presented at retrieval. Such variations may partially explain the inconsistent results found in prior color and context research. This has considerable implications for the study of context retrieval cues.

Limitations of the Study

The reader should interpret the gender x posttest color interaction results with caution. The strongly positive effects of posttest color for males and the moderately negative effects for females could be artifacts of sampling error. The male control group scored very low relative to the other treatment groups, while the female control group scored relatively high. This pattern could serve to inflate positive differences for males and negative differences for females.

One puzzling finding was the extremely poor gist memory performance of participants in this study. On average, participants barely answered one FPT question in nine correctly. Such extremely low posttest scores made it difficult to detect any context-related differences. One explanation for this *floor effect* is that, as Reyna and Kiernan (1994) suggest, the strong verbatim traces at posttest inhibited associated gist traces, thereby suppressing gist traces in general. A likelier explanation is that the encoding activity, simply reading the passage, failed to elicit sufficient mental processing to establish and reinforce strong gist traces. Strong gist traces are products of higher-level mental processes, e.g., comprehension, analysis, synthesis, and evaluation. Perhaps, if there had been encoding activities that elicited this kind of mental elaboration, recall of gist traces would have been stronger and performance on the FPT and CK items would have been higher. Another possibility is that the strength of gist memory traces simply cannot be accurately assessed through fill-in-the-blank test items. With fill-in-the-blank test items, there was a much narrower, more explicit, field of possible correct responses than there would have been with short answer and essay questions. If participants responded with gist answers (i.e., answers that were generally, if not technically, correct), the computer program would have marked their responses as incorrect. Therefore, *computerized* fill-in-the-blank test items may have a natural bias against gist responses.

Because participants were told during debriefing that color was of interest to the researchers, it is possible that some contamination may have taken place, although one would expect that this effect would occur randomly within the sample. The data collection took place over a two-week period, which only increased the chances of contamination. One way to control for this form of bias would have been to narrow the data collection window, preferably collecting all of the data in a single day or session. Unfortunately, this was not possible given the large sample size and the logistics involved.

Suggestions for Future Research

The findings support the continued study of item type and item difficulty in relation to color context effects. In future investigations, it may be beneficial to block test items into three to five groups, rather than only two, so a more precise range of activation strengths can be established for optimizing color context effects.

Future research in this area should also consider the interaction of gender effects, particularly with regard to retrieval cues. It may be useful to ask participants at the end of the experiment whether or not they noticed the presence of color and, if so, whether the presence of color was distracting, helpful, stimulating, or irritating. This control may help to explain some of the individual differences, such as the gender differences found in the current study.

On a related note, it may be beneficial to test participants' recognition of color by asking them to identify the colors seen in the course from a broader palette of hues. In the current study, participants were asked to identify topic-color associations, which is not a direct measure of color memory. Had such a measure been used, it might have been possible to detect whether or not color information was being encoded automatically.

This study examined the implicit effects of border colors, i.e., participants were not told that color would be important. In future studies, it may be of value to consider explicit context effects. If participants are told about the color-coding scheme in advance and are given an encoding activity that helps them form ensemble traces (Murnane et al., 1999), the effects of the color-matching strategy may be much stronger. On the other hand, it is possible that when learners have explicit knowledge of context information, they incur a cognitive load *tax* that offsets the benefits of the context-matching effects. Also, it is possible that strengthening context effects has negative implications for the transferability of knowledge to novel contexts. These possible effects warrant further investigation.

Lacking in the color research is a behavioral analysis of how learners interact with color in the context of CBI. One way to uncover differences in how males and females perceive and process color may be through a cognitive task analysis (Hoffman, Crandall, & Shadbolt, 1998; Hoffman & Woods, 2000). Utilizing a contextualistic approach (Foxall, 1999; Morris, 1997) would be consistent with an ecological approach to studying perception (Gibson, 1979). Both stress the interaction between the learner's mental state and environmental factors and both take a pragmatic stance that straddles the great divide between radical behaviorism and cognitivism. In any case, it seems reasonable to suggest that understanding both the affordances of color and the cognitive differences of individual learners is required in order to understand fully the contextual effects of color.

Conclusion

Contrary to encoding specificity, the matching of posttest color with lesson color, i.e., the matching of retrieval context with encoding context, did *not* optimize the conditions for recall. Border colors in the lesson had a non-significant effect (*ES* = .22) on recall. Item type, item difficulty, and gender were important factors in describing the effects of presenting border colors in the posttest. Consistent with prior gender research, females consistently recalled more to-be-remembered information than males, supporting the claim that females have an inherent advantage on constructed response test items (Weaver & Raptis, 2001).

The current investigation found that posttest color influenced recall. For males, the positive effect of posttest color is significant (*ES* = .47). For females, the negative effect is moderate but not significant (*ES* = -.36). The cause of these offsetting effects is unclear, although it appears that females noticed the border colors more, scoring higher on a topic-color matching test than males. It was posited that the collative properties of the changing border color hues in the posttest caused attention-arousal effects. It was further proposed that males effortlessly encoded the color information and therefore benefited from its attention-arousal effects without incurring a cognitive load tax. Females, however, made an effort to encode the color information, and thus incurred additional memory load. This additional load offset any attention-arousal benefits. This explanation is tentative pending further investigation of these interactions. The gender interactions should be viewed with skepticism until they are replicated with additional studies.

References

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*. New York: Academic Press.

Alderfer, C. P. (1972). *Existence, Relatedness, and Growth: Human Needs in Organizational Settings*. New York, NY: The Free Press.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete Edition)*. New York: Longman.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York, NY: Holt Rhinehart & Winston.

Backman, L., Nilsson, L. G., & Nouri, R. K. (1993). Attentional demands and recall of verbal and color information in action events. *Scandanavian Journal of Psychology, 34*, 246-254.

Baddeley, A. D. (1999). *Essentials of human memory*. East Sussex, UK: Psychology Press.

Bahrick, H. P. (1970). A two-phase model for prompted recall. *Psychological Review, 77*, 215-222.

Baine, D. (1986). *Memory and Instruction*. Englewood Cliffs, NJ: Educational Technology Publications.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122-147.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.

Beall, A. E. (1993). A social constructionist view of gender. In A. E. Beall & R. J. Sternberg (Eds.), *The psychology of gender* (pp. 127-147). New York: Guilford Press.

Bellizzi, J. A., & Hite, R. E. (1992). Environmental color, consumer feelings, and purchase likelihood. *Psychology and Marketing, 9*(5), 347-364.

Berlyne, D. E. (1960). *Conflict, Arousal, and Curiosity*. New York: McGraw-Hill.

Berlyne, D. E., Craw, M. A., Salapatek, P. H., & Lewis, J. L. (1963). Novelty, complexity, incongruity, extrinsic motivation, and the GSR. *Journal of Experimental Psychology, 66*, 476-483.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive Domain*. White Plains, NY: Longman.

Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement, 27*, 165-174.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the gist: fuzzy-trace theory and the New Intuitionism. *Developmental Review, 10*, 3-47.

Brainerd, C. J., Reyna, V. F., & Brandse, E. (1995). Are children's false memories more persistent than their true memories? *Psychological Science, 6*, 359-364.

Brannon, L. (1999). *Gender: Psychological Differences* (2nd ed.). Boston, MA: Allyn and Bacon.

Breland, H. M., Danos, D. O., Kahn, H. D., Kubota, M. Y., & Bonner, M. W. (1994). Performance versus objective testing and gender: An exploratory study of an advanced placement history examination. *Journal of Educational Measurement, 31*, 275-293.

Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement, 31*, 37-50.

Browne, K. R. (2002). *Biology at Work: Rethinking Sexual Equality*. New Brunswick, NJ: Rutgers University Press.

Canelos, J., Taylor, W. D., & Dwyer, F. M. (1985). The effects of recall cue and cognitive trace compatibility when learning from visualized instruction: An application of encoding specificity. *International Journal of Instructional Media, 12*(3), 167-178.

Chute, A. G. (1979). Analysis of the instructional functions of color and monochrome cueing in media presentations. *Educational Communication and Technology Journal, 27*(4), 251-263.

Chute, A. G. (1980). Effect of color and monochrome versions of a film on incidental and task-relevant learning. *Educational Communication and Technology Journal, 28*(1), 10-18.

Clariana, R. B. (2002). *Screen design: An interaction of color context and response mode.* Paper presented at the E-Learn 2002 Conference, Montreal, CN.

Clariana, R. B. (2003). An interaction of screen color and lesson task in CAL. *British Journal of Educational Technology*.

Clariana, R. B. (in press). The instructional effects of matching or mismatching lesson and posttest screen color. *International Journal of Instructional Media*.

Clariana, R. B., & Koul, R. (2002). The effects of different forms of feedback on fuzzy and verbatim memory of science principles. Manuscript submitted for publication.

Cronbach, L. (1963). Evaluation for course improvement. *Teachers College Record, 64*, 672-683.

Cronbach, L., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

De Mars, C. E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education, 11*, 279-299.

De Mars, C. E. (2000). Test stakes and item format interactions. *Applied Measurement in Education, 13*(55-77).

Dick, W., & Carey, L. (1990). *The Systematic Design of Instruction*. Glenview, IL: Scott, Foresman.

Dulsky, S. G. (1935). The effect of a change of background on recall and relearning. *Journal of Experimental Psychology, 18*, 725-740.

Dwyer, F. M. (1967). Adapting visual illustrations for effective learning. *Harvard Educational Review, 37*, 250-263.

Dwyer, F. M., & Lamberski, R. J. (1982-83). A review of the research on the effects of the use of color in the teaching-learning process. *International Journal of Instructional Media, 10*(4), 303-328.

Dwyer, F. M., & Moore, C. A. (1994). The effect of colour-coding and test type (visual/verbal) on students identified as possessing different field dependence levels. *British Journal of Educational Technology, 25*(3), 217-219.

Dwyer, F. M., & Moore, D. M. (2001). The effect of gender, field dependence and color-coding on student achievement of different educational objectives. *International Journal of Instructional Media, 28*(3), 309-318.

Eagly, A. H. (1987). Reporting sex differences. *American Psychologist, 42*, 756-757.

Eagly, A. H. (1997). Comparing women and men: Methods, findings, and politics. In M. R. Walsh (Ed.), *Women, Men, and Gender: Ongoing Debates* (pp. 24-31). New Haven, CT: Yale University Press.

Elio, R. E., & Reutener, D. B. (1978). Color context as a factor in encoding and as an organization device for retrieval of word lists. *The Journal of General Psychology, 99*, 223-232.

Epstein, M. L., & Dupree, D. A. (1977). Category headings as recall cues: A test of the principle of encoding specificity. *The Journal of General Psychology, 97*, 101-108.

Epstein, M. L., Dupree, D. A., & Gronikowski, L. A. (1979). Encoding specificity and contextual similarity. *Bulletin of the Psychonomic Society, 14*(3), 177-180.

Foxall, G. R. (1999). The contextual stance. *Philosophical Psychology, 12*(1), 25-46.

Freedman, D. H. (1989). Microcomputers and the dynamics of image making and social life in three art classrooms. *Journal of Research on Computing in Education, 21*, 290-298.

Gagne, R. M. (1968). Learning hierarchies. *Educational Psychologist, 6*, 1-9.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational Research: An Introduction (6th edition)*. White Plains, NY: Longman.

Gentile, D. A. (1993). Just what are sex and gender, anyway? A call for a new terminological standard. *Psychological Science, 4*, 120-122.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1977). The Theory of Affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, Acting, and Knowing*. Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1982). A preliminary description and classification of affordances. In E. S. Reed & R. Jones (Eds.), *Reasons for Realism: Selected Essays of James J. Gibson* (pp. 403-406): Lawrence Erlbaum Associates, Inc.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325-331.

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory. *British Journal of Psychology, 71*, 99-104.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23*, 553-568.

Halpern, D. F. (1994). Stereotypes, science, censorship, and the study of sex differences. *Feminism & Psychology, 4*, 523-530.

Hartley, J., & Trueman, M. (1983). The effect of headings in text on recall, search and retrieval. *British Journal of Educational Psychology, 53*, 205-214.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356-387.

Hatwell, Y. (1995). Children's memory for location and object properties in vision and haptics: Automatic or attentional processing? *Current Psychology of Cognition, 14*, 47-71.

Herzberg, F., Mauser, B., & Snyderman, B. (1959). *The Motivation to Work*. New York, NY: John Wiley.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors, 40*(2), 254-288.

Hoffman, R. R., & Woods, D. D. (2000). Studying cognitive systems in context: Preface to the special section. *Human Factors, 42*(1), 1-12.

Jacoby, L. L. (1974). The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior, 13*, 483-496.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 21-38.

Jacoby, L. L., Craik, F. I. M., & Begg, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 18*, 585-600.

James, W. (1890). *Principles of Psychology* (Vol. 1). New York, NY: Holt.

Jensen, A. R. (1969). How much can we boast IQ and scholastic achievement? *Harvard Educational Review, 39*, 81-82.

Jonassen, D. H. (1993). The trouble with learning environments. *Educational Technology, 33*(1), 35-37.

Kachigan, S. K. (1991). *Multivariate Statistical Analysis: A Conceptual Introduction* (2nd ed.). New York: Radius Press.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Keller, J. (1999). *Applying the ARCS Model of motivational design in distance learning*. Retrieved 3/31/2002, from the World Wide Web: http://mailer.fsu.edu/~jkeller/john/index.htm

Keller, J., & Kopp, T. W. (1987). An application of the ARCS Model of motivational design. In C. M. Reigeluth (Ed.), *Instructional theories in action: Lessons illustrating selected theories and models* (pp. 289-320). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kennedy, J. J., & Bush, A. J. (1985). *An Introduction to the Design and Analysis of Experiments in Behavioral Research*. Lanham, MD: University Press of America.

Keselman, H. J., Huberty, C. J., Lix, L. M., & Olejnik, S. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research, 68*(3), 350-386.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of Human Memory*. New York, NY: Academic Press.

Kirkpatrick, D. (1998). *Evaluative training programs: The four levels* (2nd ed.). New York, NY: Barrett-Kohler.

Kludt, K. T., & Dyre, B. P. (2001). Effect of color coding on misperception of heading from superimposed optic flow. *Proceedings of the Human Factors and Ergonomics Society, 2*, 1415-1419.

Kolers, P. A., Duchnicky, R. L., & Ferguson, D. C. (1981). Eye movement measurement of readability of C.R.T. displays. *Human Factors, 23*(5), 517-527.

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice, 41*(4), 212-218.

Kravitz, D. A., & Martin, B. J. (2000). *Instructor's Manual and Test Bank: Industrial and Organizational Psychology*. New York, NY: Wiley.

Lamberski, R. J., & Dwyer, F. M. (1981). The instructional effect of color in intermediate and delayed reaction. *The Journal of Instructional Psychology, 8*(4), 122-131.

Lamberski, R. J., & Dwyer, F. M. (1983). The instructional effect of coding (color and black and white) on information acquisition and retrieval. *Educational Communications and Technology Journal, 31*(1), 9-21.

Ling, J., & Blades, M. (1996). Incidental recall of color information by children and adults. *Applied Cognitive Psychology, 10*, 141-150.

Lips, H. M., Myers, A., & Colwill, N. L. (1978). Sex differences in ability. In N. L. Colwill (Ed.), *The Psychology of Sex Differences*. Englewood Cliffs, NJ: Prentice-Hall.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology, 75*(3), 402-412.

Lorch, R. F., Jr., & Lorch, E. P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology, 21*, 261-278.

Lott, B. (1997). Cataloging gender differences: Science or politics? In M. R. Walsh (Ed.), *Women, Men, and Gender: Ongoing Debates* (pp. 19-23). New Haven, CT: Yale University Press.

Maccoby, E. E., & Jacklin, C. N. (1974). *The Psychology of Sex Differences*. Stanford, CA: Stanford University Press.

Martin, E. (1975). Generation recognition theory and the encoding specificity principle. *Psychological Review, 82*, 150-153.

Maslow, A. (1954). *Motivation and Personality*. New York, NY: Harper & Row.

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics, 29*, 204-209.

Mazzeo, J., Schmitt, A. P., & Bleistein, C. A. (1993). *Sex-related performance differences on constructed-response and multiple-choice sections of advanced placement examinations* (College Board Report No. 92-7). New York: College Entrance Examination Board.

McCormack, P. D. (1972). Recognition memory: How complex a retrieval system? *Canadian Journal of Psychology, 26*, 19-41.

Merrill, M. D. (1999). Instructional transaction theory (ITT): Instructional design based on knowledge objects. In C. M. Reigeluth (Ed.), *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory* (Vol. II). Mahwah, NJ: Lawrence Erlbaum Associates.

Milheim, W. D., & Lavix, C. (1992). Screen design for computer-based training and interactive video: Practical suggestions and overall guidelines. *Performance and Instruction, 31*(5), 13-21.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miller, P. H., & Bjorklund, D. F. (1998). Contemplating fuzzy-trace theory: The gist of it. *Journal of Experimental Child Psychology, 71*, 184-193.

Moore, D. M., & Dwyer, F. M. (1997). Effect of color-coding on locus of control. *International Journal of Instructional Media, 24*(2), 145-151.

Mori, M., & Graf, P. (1996). Nonverbal local context cues explicit but not implicit memory. *Consciousness and Cognition, 5*, 91-116.

Morris, E. K. (1997). Some reflections on contextualism, mechanism, and behavior analysis. *The Psychological Record, 47*(4), 529-542.

Morris, P. (1978). Task and material variables (Chapter 2). In M. M. Gruneberg & P. Morris (Eds.), *Aspects of Memory*. London: Methuen.

Morrison, G. M., Ross, S. M., Gopalakrishnan, M., & Casey, J. (1995). The effects of feedback and incentives on achievement in computer-based instruction. *Contemporary Educational Psychology, 20*, 32-50.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 882-894.

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory and Cognition, 22*, 584-590.

Murnane, K., & Phelps, M. P. (1995). The effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 158-172.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE Theory. *Journal of Experimental Psychology: General, 128*(4), 403-415.

Murphy, R. J. L. (1982). Sex differences in objective test performance. *British Journal of Educational Psychology, 52*, 213-219.

Nitko, A. J. (1996). *Educational Assessment of Students (2nd Ed)*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

Norman, D. A. (1999). *Affordances, conventions, and design*. Retrieved 5/5/03, from the World Wide Web: http://jnd.org/dn.mss/affordances-interactions.html

Otto, W., & Askov, E. (1968). The role of color in learning and instruction. *Journal of Special Education, 2*, 155-165.

Patel, H., Blades, M., & Andrade, J. (1999). Children's incidental recall of colour information. *The British Journal of Developmental Psychology, 17*(4), 537-551.

Peck, A. C. (2000). *Industrial Psychology*. University Park, PA: Penn State University.

Pellegrino, J. W., & Salzberg, P. M. (1975). Encoding specificity in cued recall and context recognition. *Journal of Experimental Psychology: Human Learning and Memory, 104*(3), 261-270.

Pett, D., & Wilson, T. (1996). Color research and its application to the design of instructional materials. *Educational Technology Research and Development, 44*(3), 19-35.

Post, D. L., Geiselman, E. E., & Goodyear, C. D. (1999). Benefits of color coding weapons symbology for an airborne helmet-mounted display. *Proceedings of the Human Factors and Ergonomics Society, 41*(4), 515-524.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instruction, and retention interval. *Developmental Psychology, 30*, 178-191.

Richey, R. C. (1992). *Designing instruction for the adult learner: Systemic training theory and practice*. Bristol, PA: Taylor and Francis.

Rogers, P. L. (1995). *Girls like colors: boys like action? Imagery preferences and gender.* Paper presented at the National Convention of the Association for Educational Communications and Technology.

Ryder, M., & Wilson, B. (1996). *Affordances and constraints of the Internet for learning and instruction.* Paper presented at the Association for Educational Communications Technology Conference, Indianapolis, IN.

Santa, J. L., & Lamwers, L. L. (1974). Encoding specificity: Fact or artifact? *Journal of Verbal Learning and Verbal Behavior, 13*, 412-423.

Santa, J. L., & Lamwers, L. L. (1975). Where does the confusion lie?: Comments on the Wiseman and Tulving paper. *Journal of Verbal Learning and Verbal Behavior, 15*, 53-57.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 501-518.

Schwier, R. A., & Misanchuk, E. R. (1993). *Interactive Multimedia Instruction*. Englewood Cliffs, NJ: Educational Technology Systems.

Sherman, J. A. (1978). *Sex-related Cognitive Differences: An Essay on Theory and Evidence*. Springfield, IL: Charles C. Thomas.

Skinner, B. F. (1938). *The Behavior of Organisms*. New York: Appleton-Century-Crofs.

Smith, P. L., & Ragan, T. J. (1993). *Instructional Design*. New York, NY: Macmillan.

Smith, S. M. (1986). Environmental context-dependent recognition memory using a short-term memory task for input. *Memory & Cognition, 14*, 347-354.

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13-34). New York, NY: Wiley.

Smith, S. M. (1994). Theoretical principles of context-dependent memory. In P. Morris & M. Gruneberg (Eds.), *Theoretical aspects of memory* (2nd ed., pp. 168-195). New York, NY: Routledge.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*, 342-353.

Spector, P. E. (2000). *Industrial and Organizational Psychology: Research and Practice* (2nd ed.). New York, NY: Wiley.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using Multivariate Statistics* (3rd ed.). New York: Harper Collins.

Tessmer, M., & Harris, D. (1992). *Analysing the instructional setting: Environmental analysis*. Bristol, PA: Taylor and Francis.

Tessmer, M., & Richey, R. C. (1997). The role of context in learning and instructional design. *Educational Technology Research and Development, 45*(2), 85-115.

Thompson, M. E. (1994). Design considerations of visuals. In D. M. Moore & F. M. Dwyer (Eds.), *Visual Literacy: A Spectrum of Visual Learning* (pp. 165-182). Englewood Cliffs, NJ: Educational Technology Publications.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86*, 255-262.

Tulving, C. (1974). Recall and recognition of semantically encoded words. *Journal of Experimental Psychology, 102*(5), 778-787.

Tulving, C., & Watkins, M. J. (1975). Structure of Memory Traces. *Psychological Review, 82*, 262-275.

Tulving, E. (1972). Episodic and semantic memory. In C. Tulving & W. Donaldson (Eds.), *Organization in Memory*. New York: Academic Press.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory of words. *Journal of Experimental Psychology, 77*, 593-601.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.

Unger, R. K. (1979). Toward a redefinition of sex and gender. *American Psychologist, 34*, 1085-1094.

Vroom, V. H. (1964). *Work and Motivation*. New York: John Wiley and Sons.

Weaver, A. J., & Raptis, H. (2001). Gender differences in introductory atmospheric and oceanic science exams: Multiple choice versus constructed response questions. *Journal of Science Education and Technology, 10*(2), 115-126.

Wechsler, D. (1944). *The Measurement of Adult Intelligence* (3rd ed.). Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York, NY: Psych Corp.

Wechsler, D. (1958). *Measurement and Appraisal of Adult Intelligence*. Baltimore, MD: Williams & Wilkins.

Weiss, W., & Margolius, G. (1954). The effect of context stimuli on learning and retention. *Journal of Experimental Psychology, 48*, 318-322.

Wertheimer, M. (1959). *Productive Thinking* (Enlarged ed.). New York: Harper & Row.

Williams, J. A. (2000). The relative effectiveness of varied visual testing formats in retrieving information related to different educational objectives (Doctoral dissertation, The Pennsylvania State University). *Dissertation Abstracts International*(UMI No. 9998460).

Winn, W., & Everett, R. J. (1979). Affective rating of color and black-and-white pictures. *Educational Communications and Technology Journal, 27*(2), 148-156.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological Differentiation*. New York, NY: Wiley.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A Manual for the Group Embedded Figures Test*. Palo Alto, CA: Consulting Psychologists Press, Inc.

Worley, G. M., & Moore, D. M. (2001). The effects of highlight color on immediate recall on subjects of different cognitive styles. *International Journal of Instructional Media, 28*(2), 169-179.

Appendix A:

Lesson Content: Transcript of Chunked Reading Passages

Theories of Employee Motivation (Transcript)

1.0 Hierarchy of Needs

1.1 Overview

Abraham Maslow's (1943) Hierarchy of Needs Theory proposes that motivation arises from having unmet needs. Maslow grouped these needs into five categories. Click each level on the pyramid to learn more about them.

- Physiological needs - basic biological needs for things such as food, water, and sex
- Safety needs - need for personal safety, shelter, and a safe workplace
- Love needs - need for friendship, partnership, and camaraderie
- Esteem needs - need for self-respect and for the respect of others… people need to feel good about themselves
- Self-actualization needs - need for self-improvement, truth, justice, wisdom, meaning, fulfillment of personal life goals and of one's potential

1.2 Major Distinctions

Maslow's Hierarchy of Needs Theory focuses on the internal needs of the person, whereas behavioral theories focus on reactions to external stimuli.

Maslow believed that tension-reduction is the driving force behind motivation. An unmet need creates a tension to meet that need. For example, if you need food, you feel tension until the need is met. The longer you go without food, the more the need grows.

Maslow also believed that you must meet your lower needs before you can be motivated to address higher needs. As long as you are starving or fearing for your life, you will not be motivated by the need for self-improvement or fulfillment.

1.3 Implications

If motivation is fueled by unmet needs, then you should try to find out what needs are the most important for each of your workers.

Motivating your workers is about more than just money. Appealing to their need for respect, camaraderie, fulfillment, and self-improvement are just as important, if not more.

Before you can motivate your workers through higher level needs, you should make sure that basic physiological and safety needs are met.

1.4 Key Terms

- Physiological needs - basic biological needs for things such as food, water, and sex
- Safety needs - need for personal safety, shelter, and a safe workplace
- Love needs - need for friendship, partnership, and camaraderie
- Esteem needs - need for self-respect and for the respect of others… people need to feel good about themselves
- Self-actualization needs - need for self-improvement, truth, justice, wisdom, meaning, fulfillment of personal life goals and of one's potential
- Tension reduction - the driving force behind motivation… an unmet need creates a tension to meet that need

2.0 ERG Theory

2.1 Overview

Clayton Alderfer (1972) built on Maslow's work and proposed his own model for motivation. ERG Theory groups needs into three categories, varying in their degree of concreteness: existence, relatedness, and growth. Click each one to learn more about them.

- Existence needs - need for concrete, tangible things like food, water, and material possessions (equivalent to Maslow's physiological and safety needs)
- Relatedness needs - social needs and the need to have relationships with family, friends, co-workers, and supervisors (similar to Maslow's love needs)
- Growth needs - need for self-improvement or personal growth, expression of creativity and productivity

2.2 Major Distinctions

Alderfer agreed with Maslow's basic premise, that motivation is driven by internal needs. However, Alderfer proposed a new theory that more closely reflected empirical research on motivation.

Alderfer suggests that when a more concrete need is fulfilled, that need is lessened but less concrete needs become more important. In other words, when your existence and relatedness needs are satisfied, your growth needs are likely to become more important to you.

> The exception to this is that when a growth need is satisfied, that need tends to grow. If you are meeting your need to express yourself creatively, your need to express yourself further is likely to grow.

Alderfer introduced the concept of frustration-regression. Basically, when we have trouble meeting a less concrete need, we regress to meet needs at a more concrete level. For example, if you were having trouble meeting growth needs at work, you might socialize more in an effort to address relatedness needs.

> The exception to frustration-regression is that failure to meet existence needs only strengthens those needs. If you need food and can't get any, your need becomes stronger.

2.3 Implications

Workers shift back and forth from more concrete to less concrete need types depending on their individual situations.

When more concrete existence and relatedness needs are satisfied, workers are more likely to focus on growth needs.

Providing your workers with growth opportunities can create the need for more and more growth opportunities.

If your workers are not given ample opportunity for growth, they can get frustrated and regress toward more concrete needs, such as socializing.

2.4 Key Terms

- Existence needs - need for concrete, tangible things like food, water, and material possessions (equivalent to Maslow's physiological and safety needs)
- Relatedness needs - social needs and the need to have relationships with family, friends, co-workers, and supervisors (similar to Maslow's love needs)
- Growth needs - need for self-improvement or personal growth, expression of creativity and productivity
- Frustration regression – the idea that when we fail to meet a growth or relatedness need, we refocus our efforts on more concrete needs

3.0 Two-Factor Theory

3.1 Overview

In his Two-Factor Theory of motivation, Frederick Herzberg (1959) argues that there are two types of factors involved in motivation: extrinsic and intrinsic. Click on each one to learn more.

> Extrinsic (or hygiene) factors include tangible outcomes and things that focus on a worker's physical well-being such as pay and benefits, organizational policies, quality of supervision, job security, job safety, administrative practices, and physical work conditions. These factors do not lead to motivation, but without them there can be dissatisfaction.
>
> Intrinsic (or motivator) factors include intangible outcomes such as achievement, recognition, responsibility, growth, respect, and interest in the job. According to Herzberg, only intrinsic factors motivate.

3.2 Main Points

Herzberg basically argues that extrinsic factors do not motivate people. You are not motivated by having fair policies at work, although you might be highly dissatisfied if a policy were changed in a way that you thought was unfair. The best you can hope for with extrinsic factors is to keep workers from feeling dissatisfied.

Intrinsic factors, on the other hand, are the things that can make us feel good about going to work in the morning. If you feel respected, enjoy the level of responsibility, and get a kick out of the work itself, then you are likely going to be satisfied with your job. In the absence of motivator factors, you will (at best) feel no satisfaction.

Herzberg's primary contribution was to shift our focus away from tangible factors (such as money) to intangible factors such as praise, responsibility, and recognition.

Based on the Two-Factor Theory, Herzberg created the Orthodox Job Enrichment (OJE) model. OJE is based on the idea that organizations can elevate motivation by re-designing jobs to improve intrinsic motivation.

3.3 Implications

In order to motivate workers, you should focus on both the extrinsic and intrinsic factors.

Focusing all of your attention on extrinsic factors will yield at best a worker who is not dissatisfied.

Re-designing jobs so that workers have more responsibility, freedom to make decisions, and more challenging work is one way to improve satisfaction levels.

Workers who really gain intrinsic satisfaction can perform at high levels even when extrinsic factors are low.

3.4 Key Terms

Extrinsic (Hygiene) Factors – include tangible outcomes and things that focus on a worker's physical well-being.

Intrinsic (Motivator) Factors - include intangible outcomes such as recognition, responsibility, respect, and interest in the job.

OJE (Orthodox Job Enrichment) – A job redesign program developed by Herzberg that emphasizes intrinsic factors.

4.0 Reinforcement Theory

4.1 Overview

Reinforcement Theory is the product of B.F. Skinner's (1938) behavioral studies of operant conditioning, which emphasized the importance of rewards and punishment in stimulating desired behaviors.

The central premise of reinforcement theory is that the consequences (or outcomes) of behavior influence the likelihood that people will behave the same way again. Behavior can be influenced in three ways: through positive reinforcement, negative reinforcement, and punishment. Click on the illustration below for a description of each:

Positive reinforcement
This is a form of *reward* that involves giving something that is liked or wanted as a consequence of some behavior. For example, high performing workers sometimes receive bonuses, praise, or recognition.

Negative reinforcement
This is a form of *reward* that involves removing something that is disliked as a consequence of some behavior. For example, a worker might stop coming to work late in order to avoid being lectured again by his/her supervisor.

Punishment
People can be punished for doing something inappropriate by the removal of something they like or by the addition of something they dislike.

4.2 Main Points

Reinforcement Theory focuses on behavior and treats it as simply a reaction to external forces. This is different from cognitive theories of motivation, which focus on internal states, and need theories, which focus on internal needs.

Reinforcement Theory recognizes that rewards can be performance-oriented or time-oriented. Since both have their strengths and weaknesses, most standard compensation systems provide time-based *and* performance-based rewards. Click on the illustration below for a description of each:

Time-based rewards

Interval schedules reward correct behaviors based on a time interval. Organizations often adopt a weekly, bi-weekly, or monthly pay schedule and set up bonus and promotion opportunities based on seniority.

Although time-based reinforcement schedules reward people for the time they spend working, they do not capture the quantity or quality of the work performance. In the worst scenarios, lazy workers are rewarded for their ability to avoid termination rather than their performance.

Performance-oriented rewards

Continuous reinforcement schedules reward people after every instance of a correct behavior. For example, people who work on commission are rewarded after each successful sale.

Ratio schedules reward people for producing output of a certain quantity or quality. For example, an employee of the month receives a bonus and a special parking space as rewards for good performance. A production line worker who is on a *piece rate* plan is paid a set price for each item he/she successfully produces.

Performance-oriented schedules reward the amount or quality of work output without considering the time spent to meet the quota. This can cause morale problems, since workers can spend a lot of time performing well without getting a reward. A good sales person may go weeks or months without selling anything and a worker may perform well for years without being selected as employee of the month.

4.3 Implications

Reinforcement theory compels us to be more sensitive to how intentional and unintentional rewards and punishments impact worker behavior.

Environmental influences can have a significant impact on a worker's behavior.

Managers can give or withhold praise and other rewards in a manner that will shape the behavior of their workers in ways that improve specific behaviors.

The frequency with which rewards are given can impact performance. Continuous, ratio, and interval reinforcement schedules can produce differing results.

4.4 Key Terms

Positive reinforcement - a form of *reward* that involves giving something that is liked or wanted as a consequence of some behavior.

Negative reinforcement - a form of *reward* that involves removing something that is disliked as a consequence of some behavior.

Punishment - People can be punished for doing something inappropriate by the removal of something they like or by the addition of something they dislike.

Continuous reinforcement - Rewards people after every instance of a correct behavior.

Ratio schedules - Reward people for producing output of a certain quantity or quality.

Interval schedules - Reward correct behaviors based on a time interval.

5.0 Expectancy Theory

5.1 Overview

Whereas Skinner's Reinforcement Theory explains how different types of reinforcement shape behavior, Victor Vroom's (1964) Expectancy Theory helps predict to what degree a particular reward will impact an individual's motivation level.

According to Expectancy Theory, motivation is a function of an individual's confidence that he/she can perform a behavior successfully (expectancy), that performing successfully will lead to an outcome (instrumentality), and that the outcome is desirable (valence). Only if all three elements are high can motivation be high.

5.2 Main Points

According to Vroom, motivation can be expressed as a mathematical equation: $F = E \times V \times I$. Click each part of the formula below to learn more:

The Motivation Formula

$$F = E \times V \times I$$

Force is the amount of motivation a person has to engage in a particular behavior (e.g., motivation to be highly productive at work).

Expectancy is the confidence an individual feels that he/she can perform the behavior successfully. This is normally stated as a probability (e.g., 80% confident that I can be highly productive).

Valance is the value a person assigns to that outcome (e.g., a raise would be 70% desirable but a promotion would be 95% desirable).

Instrumentality is the confidence an individual feels that performing the behavior will result in a particular outcome. Again this is expressed as a probability (e.g., 80% sure that high productivity will lead to a raise or promotion).

5.3 Implications

Different rewards motivate different people to different degrees.

> Consider, for example, two sales people who are eligible for a promotion. For one, the possibility of being promoted is very attractive and so his valence is high (+ 90%), but given his prior performance his expectancy is low (10%). For the other sales person, the promotion is moderately attractive (valance = + 70%) and she is moderately confident that she can get the promotion (expectancy = 70%). Assuming that both have equal instrumentalities (of say 100%), the first sales person has a motivation force of .09 (.9x1x.1) while the second has a force of .49 (.7x1x.7). This approach emphasizes the individual nature of motivation.

The Motivation Formula suggests that motivation (force) cannot exist unless the individual possesses at least some expectancy, instrumentality, *and* valance. If any one of them is zero, there is no motivation.

Expectancy Theory suggests that you should get to know what your workers value and then work to improve the confidence they have in their ability to gain the rewards that they value.

5.4 Key Terms

Force - the amount of motivation a person has to engage in a particular behavior.

Expectancy - the confidence an individual feels that he/she can perform the behavior successfully.

Instrumentality - the confidence an individual feels that performing the behavior will result in a particular outcome.

Valance - the value a person assigns to an outcome (reward).

6.0 Empowerment Theory

6.1 Overview

Empowerment theory is an extension of Albert Bandura's (1982) Self-efficacy Theory. Bandura suggested that our motivation and performance are in part dependent on our self-efficacy, the degree to which we believe we can accomplish the task. Motivation to perform a task increases when we have high self-efficacy towards that task.

According to empowerment theory, feelings of self-efficacy come from having competence and self-determination.

6.2 Main Points

Empowerment Theory has been used widely in organizational settings. Research has consistently found that self-efficacy predicts performance.

- Prior experience - Self-efficacy may develop from prior good performance. Similarly, low self-efficacy can be a product of previous failures.

- Challenges - People with higher self-efficacy tend to set more challenging goals.

- Commitment – People with higher self-efficacy tend to exhibit greater commitment to their goals, and so are able to overcome obstacles more easily.

- Effort - Those who have high self-efficacy are more likely to try hard, which makes them more likely to succeed. Those who have low levels of self-efficacy feel that they are not good at the task and may not try very hard at all, which makes them less likely to succeed. This is known as the Galatea Effect.

6.3 Implications

Sometimes organizations empower employees by asking them to participate in making organizational decisions. Research has shown that participatory decision-making can increase commitment to the decision that is made and improve motivation.

Flextime is a program that allows workers to design their own work schedule, empowering them to structure their workday, within limits. Research has shown that job performance and job satisfaction do benefit from flextime programs, but only sometimes. The most reliable benefit seems to be reduced absenteeism, since the program allows flexibility to manage midday appointments and reduces work-family conflicts.

Since self-efficacy can be affected by prior experiences, workers should be given new responsibilities gradually. Empowering workers who do not feel competent to perform their new tasks can be harmful.

Training, mentorship, and apprenticeships can help to develop the competence and confidence workers need to perform new responsibilities.

6.4 Key Terms

Self-efficacy - degree to which we believe we can accomplish the task.

Galatea Effect - those who have high self-efficacy, try harder and more often succeed, while those who have low self-efficacy put forth less effort and therefore usually fail.

Commitment – ability to overcome obstacles in the pursuit of a goal.

Participatory decision-making - asking workers to participate in making organizational decisions in order to increase commitment to the decision and improve motivation.

Flextime - a program that allows workers to design their own work schedule, within limits.

7.0 Equity Theory

7.1 Overview

J. Stacey Adams' (1965) Equity Theory (a.k.a., Social Exchange Theory) suggested that our effort depends on our perceptions of fairness. Adams said that as we act to satisfy our needs, we each assess the fairness of the outcome by asking ourselves: "Am I getting what I deserve in this exchange?"

According to this theory, people compare their equity ratios to those of their peers. When the ratio reflects an inequity, tension is created and so people work to reduce that tension.

In figuring out what is equitable and what is not, we compare the ratio of our outputs to inputs with the ratio of others around us. Click the illustration for an explanation of inputs and outputs.

- Inputs are the things you give to your employer - your qualifications, past experiences, seniority, results, loyalty, time, effort, etc. - essentially, what you put into your job.

- Outputs are the things you receive in return – salary, bonuses, benefits, appreciation, respect, recognition, stability, etc. – basically, what you get out of your job.

7.2 Main Points

According to Adams, we are motivated to achieve a condition of equity in our dealings with others. When we find ourselves in inequitable situations, we experience feelings of emotional tension. This inequity tension motivates us to correct the injustice.

According to equity theory, inequity tension can work both ways. When people feel that they give more and get less in return than their co-workers, they feel resentment. Similarly, when people feel that they get more than their peers, they feel guilt.

To reduce feelings of resentment due to underpayment, people are motivated to steal or in some way increase outputs to make up the difference. They may also attempt to sabotage an "overpaid" worker or the organization.

Similarly, to avoid feelings of guilt due to overpayment, people are motivated to work harder or somehow increase their inputs to even the score. In both cases, workers may resign because of the mounting resentment or guilt.

7.3 Implications

Adams reminds us of the importance of justice and fairness in the workplace.

It is important for managers to try to understand how their workers perceive fairness

Research has shown that people are motivated to act when they feel cheated. Less research supports the idea that people are motivated to act when they are overpaid in some way.

While research has found that employee perceptions of inequity correlates with intentions to quit and job search behavior, it is often difficult to tell what workers will perceive as inequitable and how they will respond to inequities. It may vary by individual, even within a given context.

7.4 Key Terms

Inputs - the things you give to your employer - your qualifications, past experiences, seniority, results, loyalty, time, effort, etc. - essentially, what you put into your job.

Outputs - the things you receive in return – salary, bonuses, benefits, appreciation, respect, recognition, stability, etc. – basically, what you get out of your job.

Equity Ratio - the ratio of outputs to inputs: the total perceived value of our outputs divided by the total perceived value of our inputs for a given situation.

Inequity Tension – The feelings of resentment or guilt that arise from situations in which our equity ratio is below that of others.

8.0 Goal-Setting Theory

8.1 Overview

Building on Bandura's self-efficacy research, Edwin Locke and Gary Latham (1990) proposed Goal-setting Theory. According to Goal-setting Theory, goals direct our mental and physical actions. Goals serve two functions:

- Goals serve as performance targets that we strive to reach.
- Goals serve as standards against which we measure our own performance.

Locke and Latham argue that the outcome of your performance can affect your future effort. In this way, goals provide you with a means of regulating your own effort.

8.2 Main Points

According to Locke and Latham, behavior is motivated by internal goals, objectives, or intentions.

Workers are motivated to accomplish specific, difficult goals assuming performance feedback is provided regularly and workers are committed to the goal.

- Goal specificity - Research has shown that workers who have vague goals are more likely to be satisfied with average performance even though they are capable of better performance. Specific goals provide workers with a sharper point of focus and are more likely to improve motivation and performance.
- Goal difficulty - Workers tend to give more effort when they are trying to reach harder goals.
- Performance feedback - When comparing performance against goals, performance feedback is generated. This feedback enables workers to regulate their efforts.
- Goal Commitment – The degree to which individual workers accept their goals. Goals can only be motivating if workers are personally committed to them.

A practical extension of goal-setting theory is the Management by Objectives (MBO) strategy. With MBO, the manager and the worker meet and agree on performance goals, which are derived from higher-level organizational goals. These are used to evaluate the worker's performance later.

8.3 Implications

The success of MBO strategies suggests that giving workers an active voice in deciding what their goals should be increases goal commitment.

The most effective goals are specific and difficult. When setting goals, encourage your workers to specify criteria for goal achievement and to challenge themselves.

When setting goals, also consider how you will generate frequent performance feedback.

Under the right conditions, failure to meet a goal can motivate workers to work harder.

You should ensure that your workers have the competencies and resources to accomplish their goals. Otherwise, failure could lead to frustration.

8.4 Key Terms

Goal specificity – The degree to which a goal statement provides specific criteria for success. For example, the goal "raise profitability 10% this year" is more specific than "Let's be more profitable." More specific goals are generally more motivating.

Goal difficulty – The degree to which a worker considers a goal challenging. More difficult goals are generally more motivating.

Performance feedback – A comparison of performance against goals. Providing frequent performance feedback sustains the motivation effects of goals

Goal Commitment – The degree to which individual workers accept their goals.

Competencies – The knowledge, skills, and attitudes (KSAs) that comprise a worker's capacity to accomplish a goal.

Management-by-objectives (MBO) – A popular management approach in which workers and their supervisors collaborate with each other to set individual goals that align with higher-level organizational goals

Appendix B:

Prior Knowledge Test: General Knowledge of Industrial and Organizational Psychology (Test Item List)

Learning Assessment Quiz

1) Organizational psychology is concerned with _____
   a) Understanding behavior and enhancing the well being of employees in the workplace.
   b) Efficient job design, performance appraisal, employee selection, and training
   c) Organizational efficiency
   d) Studying the structure of organizations
2) Who developed the application form?
   a) Hugo Munsterberg
   b) Walter Dill Scott
   c) Frederick Winslow Taylor
   d) Robert Yerkes
3) The two fathers of I/O psychology are:
   a) William James and Bruce V. Moore
   b) Hugo Munsterberg and Frederick Winslow Taylor
   c) Walter Dill Scott and Hugo Munsterberg
   d) Walter Dill Scott and Frederick Winslow Taylor
4) _____ developed Scientific Management principles as an approach to handling production workers in factories:
   a) Hugo Munsterberg
   b) Walter Dill Scott
   c) Frederick Winslow Taylor
   d) Robert Yerkes
5) The Civil Rights Act of 1964 made it illegal to make employee status decisions based on _____.
   a) Physical disability
   b) National origin
   c) Marital status
   d) Mental disability

6) I argued that basic experimental work is needed to establish psychology as a legitimate science. I argued that applied psychology couldn't contribute new knowledge to science. Who am I?

a) Hugo Munsterberg

b) Wilhelm Wundt

c) James McKeen Cattell

d) Frederick Winslow Taylor

7) The US Army began to use I/O psychologists in _____, in the first large-scale application of psychological testing to place individuals into jobs.

a) World War I

b) The 1920s

c) The 1930s

d) World War II

8) I/O psychology is concerned with _____

a) The study and treatment of emotional problems in the workplace.

b) The development and application of scientific principles to the workplace

c) The development and application of scientific principles into the workplace for the benefit of management

d) The psychological study of idiots (I) and oddballs (O)

9) Industrial psychology is concerned with _____

a) understanding behavior and enhancing the well-being of employees

b) efficient job design, performance appraisal, employee selection, and training

c) studying the structure of organizations

d) psychoanalyzing sick industries (e.g., the US automobile industry in the 70s)

10) Which of the following is NOT a principle of Scientific Management?

a) Each job should be analyzed to determine the best way of doing it.

b) Employees should be carefully trained at their jobs

c) Employees should be rewarded for productivity

d) Employees should be selected so as to get along with other members of their group

11) In a well-known Hawthorne study, employees were taken to a special room where lighting levels were varied from day to day to see how performance would be affected. They found that _____

a) Performance depended not only on lighting level, but also on whom an employee was working with that day.

b) Newer employees did better with more light, while more experienced employees were unaffected by light conditions.

c) No matter how light levels were changed, productivity remained the same unless established work groups were broken up.

d) Over the course of the study, productivity increased and seemed to have little to do with lighting levels.

12) Why was the Army Beta test created?

a) The Army Alpha test was a failure

b) The Army Alpha test was too easy

c) Many soldiers were illiterate, and therefore unable to perform well

d) They liked the name Army Beta test more than the Army Alpha test

13) Who received the first Ph.D. in what was then called Industrial Psychology?

a) James McKeen Cattell

b) Bruce V. Moore

c) Hugo Munsterberg

d) William James

14) The American's with Disabilities Act _____

a) Established a glass ceiling

b) Made it illegal to discriminate against only Americans with physical disabilities

c) Tried to decrease the unemployment rate of qualified disabled people

d) Only refers to psychological disorders

15) In 1991, the Department of Labor stated that there is a glass ceiling in many companies and corporations. What is meant by the term "glass ceiling?"

a) The idea that the only way to get management level job is to know other upper level management people

b) It is the idea that minorities and women are blocked from management-level positions

c) The idea that only people with certain types of personalities can be offered management level jobs

d) None of the above

16) I was Wundt's grad student. I believed that applied psychology could generate new knowledge and profit. Who am I?

a) Hugo Munsterberg

b) Walter Dill Scott

c) Frederick Winslow Taylor

d) James McKeen Cattell

17) I studied individual differences and developed the Psychological Corporation. Who am I?

a) Hugo Munsterberg

b) Walter Dill Scott

c) Frederick Winslow Taylor

d) James McKeen Cattell

Appendix C:

Posttest (Test Item List)

	Responses	
1. According to Maslow, _____ needs are the basic biological needs for things such as food, water, and sex.	physiological	1.1 Lesson: Hierarchy of Needs Type: FVE
2. According to Maslow, _____ needs are needs for personal safety, shelter, and a safe workplace.	safety	1.2 Lesson: Hierarchy of Needs Type: FVE
3. According to Maslow, _____ needs are needs for friendship, partnership, and camaraderie.	love	1.3 Lesson: Hierarchy of Needs Type: FVE
4. According to Maslow, _____ needs are needs for self-respect and for the respect of others.	esteem	1.4 Lesson: Hierarchy of Needs Type: FVE
5. According to Maslow, needs for personal growth and development are categorized as _____ needs.	self-actualization	1.5 Lesson: Hierarchy of Needs Type: FPTE
6. Maslow suggests that motivation is the product of a phenomenon known as _____ (2 words). This fuels our natural drive to resolve unmet needs.	tension reduction	1.6 Lesson: Hierarchy of Needs Type: FPTH
7. According to Alderfer, the needs for shelter and for money are categorized as _____ needs.	existence	2.1 Lesson: ERG Type: FPTH
8. According to Alderfer, making friends at work and spending quality time with family are expressions of _____ needs.	relatedness	2.2 Lesson: ERG Type: FPTE
9. According to Alderfer, _____ needs are needs for self-improvement, personal growth, creativity, and productivity.	growth	2.3 Lesson: ERG Type: FVE
10. According to Alderfer, when we are unable to meet an abstract need (e.g., creativity), we then focus our efforts on more concrete needs (e.g., socializing or making money). He calls this phenomenon _____ (2 words).	frustration regression	2.4 Lesson: ERG Type: FPTH

11. According to Herzberg, _____ factors are tangible outcomes and things that focus on a worker's physical well-being.	Extrinsic hygiene	3.1 Lesson: Two-Factor Type: FVE
12. According to Herzberg, _____ factors are intangible outcomes such as recognition, responsibility, respect, and interest in the job.	Intrinsic motivator	3.2 Lesson: Two-Factor Type: FVE
13. The program developed by Herzberg that involves redesigning jobs and stresses the importance of motivating outcomes is called _____ (acronym is acceptable).	orthodox job enrichment oje	3.3 Lesson: Two-Factor Type: FPTE
14. According to Skinner, _____ (2 words) is a form of reward that involves giving something that is liked or wanted as a consequence of some behavior.	positive reinforcement	4.1 Lesson: Reinforcement Type: FVE
15. According to Skinner, _____ (2 words) is a form of reward that involves removing something that is disliked as a consequence of some behavior.	negative reinforcement	4.2 Lesson: Reinforcement Type: FVE
16. According to Skinner, _____ is the removal of something a person likes or the addition of something a person dislikes.	punishment	4.3 Lesson: Reinforcement Type: FVE
17. A commission-only sales person who is rewarded after each sale of a product is an example of someone who is on a(an) _____ schedule (2 words).	continuous reinforcement	4.4 Lesson: Reinforcement Type: FPTE
18. A contractor who is paid only upon completion of work related to a project is an example of someone who is on a(an) _____ schedule.	ratio	4.5 Lesson: Reinforcement Type: FPTE
19. An office worker who is paid the same amount every two weeks is an example of someone who is on a(an) _____ schedule.	interval	4.6 Lesson: Reinforcement Type: FPTE
20. According to Vroom, _____ is the amount of motivation a person has to engage in a particular behavior.	force	5.1 Lesson: Expectancy Type: FVH

behavior.		
21. According to Vroom, _____ is the confidence an individual feels that he/she can perform the behavior successfully.	expectancy	5.2 Lesson: Expectancy Type: FVH
22. According to Vroom, it is important that a person feel that performing a task will lead to particular reward. He called this feeling _____.	instrumentality	5.3 Lesson: Expectancy Type: FPTH
23. Each of us assigns different values to possible rewards or outcomes. Vroom refers to this value as a person's _____.	valance	5.4 Lesson: Expectancy Type: FPTE
24. According to Bandura, the _____ (2 words) means that those who have high self-efficacy, try harder and more often succeed, while those who have low self-efficacy put forth less effort and therefore usually fail.	galatea effect	6.2 Lesson: Empowerment Type: FVH
25. According to Bandura, _____ is the ability to overcome obstacles in the pursuit of a goal.	commitment	6.3 Lesson: Empowerment Type: FVH
26. Conducting focus groups to gather employee input on important organizational decisions is an example of _____ decision-making.	participatory	6.4 Lesson: Empowerment Type: FPTH
27. _____ is a program that allows workers to design their own work schedule.	flextime	6.5 Lesson: Empowerment Type: FVH
28. According to Adams, the tangible and intangible things that you put into your job are called _____.	inputs	7.1 Lesson: Equity Type: FPTH
29. According to Adams, the tangible and intangible things that you get out of your job are called _____.	outputs	7.2 Lesson: Equity Type: FPTE
30. When you take the total value of what you put into your job and divide it by the total value of what you get out of your job, you come up with a(an) _____ (2 words).	equity ratio	7.3 Lesson: Equity Type: FPTH

31. According to Adams, the feelings of resentment or guilt that arise from comparing our situations with those of others are called _____ (2 words).	inequity tension	7.4 Lesson: Equity Type: FPTH
32. According to Locke and Latham, the extent to which a goal statement details the criteria for success is called goal _____.	specificity	8.1 Lesson: Goal Setting Type: FPTH
33. According to Locke and Latham, goal _____ is the degree to which a worker considers a goal challenging.	difficulty	8.2 Lesson: Goal Setting Type: FVE
34. According to Locke and Latham, goals will continue to be motivating only if people are provided frequent ______ (2 words).	performance feedback	8.3 Lesson: Goal Setting Type: FPTH
35. _____ are the knowledge, skills, and attitudes (KSAs) that comprise a worker's capacity to accomplish a goal.	competencies	8.5 Lesson: Goal Setting Type: FVH
36. _____ (acronym is acceptable) is a popular management approach in which workers and their supervisors collaborate with each other to set individual goals that align with higher-level organizational goals.	management-by-objectives mbo	8.6 Lesson: Goal Setting Type: FVH
37. Maslow proposed a model with five categories of needs. This model comes from _____ Theory.	hierarchy of needs	9.1 Lesson: N/A Type: CKH
38. Alderfer proposed a model consisting of three categories of needs. This is a product of _____ Theory.	erg	9.2 Lesson: N/A Type: CKH
39. The distinction between intrinsic and extrinsic factors is most closely associated with _____ Theory.	two-factor	9.3 Lesson: N/A Type: CKH
40. The use of rewards and punishments is most consistent with _____ Theory.	reinforcement	9.4 Lesson: N/A Type: CKH

41. The use of a mathematical formula to predict motivation levels is most consistent with _____ Theory.	expectancy	9.5 Lesson: N/A Type: CKE
42. Promoting the importance of self-efficacy is the main contribution of _____ Theory.	empowerment	9.6 Lesson: N/A Type: CKE
43. Worker perceptions of inputs and outputs are most closely associated with _____ Theory.	equity	9.7 Lesson: N/A Type: CKE
44. Management by objectives is an approach that is most closely associated with _____ Theory.	goal-setting	9.8 Lesson: N/A Type: CKE

Appendix D:

Pilot Study Design and Results

Pilot Design

A 2 x 2 x 2 experimental design was used for this study, with retrieval cueing strategy (context-matching versus context-mismatching), item format (multiple choice versus constructed response), and gender as the independent variables. The first two are between group variables, while the third is a within participants variable. The dependent variable was posttest achievement. Participants were randomly assigned to one of four treatment conditions.

Pilot Participants and Settings

Participants were 44 undergraduate students taking a 100-level *Introduction to Psychology* course at a major northeastern research university in a summer semester. Students received course credit for participating in the study. Most of the participants were freshman (.66), 18 years old (.68), undeclared (.39), and female (.61). The average SAT verbal score (self-reported) was 576, ranging from 420 to 750. There were only four declared Psychology majors in the sample, three of whom were freshmen. This is relevant since the subject matter selected for the treatments was derived from a 200-level Psychology course.

The experiment was conducted in two settings. Approximately half of the participants (.52) went through the experiment in one of four sessions held in a computer lab setting. The other half (.48) was scheduled by appointment to undergo the experiment on one of two computers setup in the office of one of the researchers. In both settings, the experiment was proctored and any unusual events were logged.

Pilot Design and Materials

This experimental study used a four-group posttest only design to determine differences in learning achievement. The dependent variable was achievement posttest scores on a test of forced-choice recognition and a test of cued recall. The main independent variable examined was the retrieval cueing strategy in the posttest (color context matching or mismatching). This design yielded four experimental groups.

All four groups read a 4,000-word computer-based tutorial – developed using Macromedia Authorware 5.2 and Adobe Photoshop 6.0 – dealing with the theories of achievement motivation (Peck, 2000). The content was rated at the 12th grade reading level, according to the Flesch-Kincaid scale. The interactive tutorial was chunked into 5 lessons, each dealt with a different group of motivation theories: (1) need, (2) behavioral, (3) self-efficacy, (4) equity, and (5) goal-setting theories. It was hoped that by using subject matter that was relevant and yet unfamiliar to the participants, student learning motivation would be high. This strategy is supported by Morrison, Ross, Gopalakrishnan, and Casey's (1995) study – involving graduate-level *Instructional Systems Design* and *Educational Psychology* classes – in which the researchers observed a much larger effect size in the class given content that was relevant to their field of study.

The courseware utilized a combination of text, still photographs, and diagrams, but no audio, video, or animation. Its screen design incorporated common web page design strategies, such as a left-hand vertical navigation bar, rollover buttons, and a header-navigation bar-content frameset layout. Unlike typical web-based courses, however, the navigation was restricted to linear branching primarily for the purpose of controlling for sequence.

After reading through each lesson, participants answered a series of four or five practice questions (the item format matched the treatment condition) and were given feedback (were told if their answers were right or wrong and were given the correct answers). The purpose of the practice questions was to reinforce the key points in each lesson. There were 22 practice items in all and each item was paraphrased from 22 corresponding items on the posttest.

Pilot Criterion measures

The achievement posttest consisted of 30 questions, 22 that were alternate versions of the ones used in the practice questions and 8 that were new questions designed to test the learner's ability to discriminate among the 5 sets of motivation theories covered in the tutorial. There were two posttests developed: a fill-in test to measure cued recall and a multiple choice test to measure recognition. Both tests were scored electronically by the courseware, but the fill-in test was also scored manually to account for any unanticipated spelling errors. The spelling errors were distributed randomly and so did not impact the main effects. An item analysis revealed that the fill-in test had a difficulty level of .43 and a Cronbach Alpha of .77 (after manual scoring), while the multiple-choice test had a difficulty level of .69 and a Cronbach Alpha of .57 (Cronbach, 1963; Nitko, 1996).

In addition to posttest achievement scores, data were collected through the computer program for demographics, student satisfaction with the course, performance on a color-blindness test, and achievement on a color-topic matching task.

The demographic information was collected with a survey that asked about gender, major, year, SAT scores, prior experience with online courses, and prior knowledge of motivation theories. In a heart content study involving 126 female and 53 male undergraduate students, Dwyer and Moore (2001) found that "a significant interaction existed between gender and color" (p.313). The criterion posttest scores of female students appeared to be more sensitive to the effects of color-coding than male students. Other studies suggest that females are more attuned to color than males (Freedman, 1989; Rogers, 1995).

Student satisfaction was gauged with a standard five-point Likert-scale questionnaire (5 being the highest score) that asked about the clarity of the directions, how well key points were presented, and overall student satisfaction with the course. Comments about what students liked and suggestions for improvement were requested also. Most participants rated as average or better the presentation of key points (.84), the clarity of the instructions (.96), and their overall satisfaction (.68) with the course.

A color-blindness test—developed by Shinobu Ishikara at the University of Tokyo (available at *http://www.copresco.com/links/colblind.htm*) and commonly administered by eye doctors – was administered to ensure that all participants were able to perceive the color context manipulated in the experiment. The test was given after the course was completed to avoid cueing the learner in on the importance of color in the experiment. All participants scored perfectly on this test.

As a check to see if participants were able to retain and recognize the association between color and lesson topic that was established in the course, a topic-color matching test was administered.

Pilot Procedures

Upon arrival, participants were given an index card containing a subject identification number and a black-and-white image. The image corresponded to one of the four treatment conditions (e.g., triangle = condition 1, rectangle = condition 2, etc.). The index cards had already been randomly ordered in sets of four and each set had been randomly ordered as well.

After signing their informed consent forms, participants were taken to a computer terminal and asked to complete the course. A proctor was available to monitor the experiment and address any technical problems. The experiment lasted one hour.

There were four treatment groups (see Table 1). All of them went through the same lesson content and all of them completed practice exercises at the end of each lesson. Each of the five lessons was assigned a particular background color: blue, green, orange, purple, and red. This color appeared in the navigation bar, along the borders of the screen, and in a swirl that ran down the right side of the screen. When the participants completed all of the lessons, they completed one of two 30-item posttests. One test consisted of forced multiple-choice (MC) items (4 foils) and the other consisted of constructed response (CR), or fill-in-the-blank, items. The posttests varied in another way: the background color used. In the matching, or condition, the same colors used in the lesson were used in the posttest to cue the correct response. In the mismatching condition, a different color was used in an attempt to mis-cue the participant.

Table 1

Overview of the study design

		Retrieval Cueing Strategy	
		Context matching	Context mismatching
Item Format	Multiple Choice	Group 1	Group 2
	Constructed Response	Group 3	Group 4

Pilot Results

Main context effects

A 2 x 2 analysis of variance (ANOVA) revealed posttest score differences favoring the mismatching conditions for both item formats (see Table 2), but the differences were not significant. In the recognition condition (multiple choice test items), the mean posttest score was 19.82 (66%) for participants in the matching group and 21.64 (72%) for those in the mismatching group. The difference (1.82) was not statistically significant at the .05 level, F (1, 20) = 1.59, p = .221. In the recall condition (constructed response test items), the mean posttest score was 12.10 (40%) for the matching group and 13.67 (46%) for the mismatching group. Again, the difference (1.57) was not statistically significant, F (1, 20) = .56, p = .463.

Table 2:

Posttest achievement by treatment condition

	Cueing Strategy Matching	Mismatching	Difference	F-Value	P-Value
Recognition					
M	19.82	21.64	(1.82)	1.59	0.221
SD	2.93	3.78			
n	11	11			
Recall					
M	12.10	13.67	(1.57)	0.56	0.463
SD	4.23	5.37			
n	10	12			

Note: Maximum score = 30

Context effects by gender

A 2 x 2 analysis of variance (ANOVA) revealed posttest score differences favoring male students for both item formats (see Table 3), but the differences were not significant. On the recognition task, the mean posttest score was 20.15 (67%) for female participants and 21.56 (72%) for male ones. The difference (1.40) was not statistically significant at the .05 level, F (1, 20) = .89, p = .358. In the recall condition, the mean posttest score was 12.14 (40%) for females and 14.38 (48%) for males. Again, the difference (2.23) was not statistically significant, F (1, 20) = 1.09, p = .309.

Table 3

Posttest achievement by gender

	Gender				
	Female	Male	Difference	F-Value	P-Value
Recognition					
M	20.15	21.56	(1.40)	0.89	0.358
SD	4.06	2.19			
n	13	9			
Recall					
M	12.14	14.38	(2.23)	1.09	0.309
SD	4.59	5.24			
n	14	8			

Note: Maximum score = 30

A 2 x 2 multivariate analysis of variance (MANOVA) revealed no significant interactions for recognition task performance between gender, λ (1, 19) = 1.173, *p* = .292, and cueing strategy, λ (1, 19) = 1.856, *p* = .189, or for cued recall task performance between gender, λ (1, 19) = .892, *p* = .357, and cueing strategy, λ (1, 19) = .393, *p* = .538.

Achievement by setting

Since two different experimental settings were used in this study, mean scores for both settings were analyzed. A MANOVA revealed an insignificant interaction between posttest performance and experimental setting across treatment conditions, λ (1, 19) = .434, *p* = .518.

Topic-color matching test

An item analysis of the topic-color matching test revealed that very few participants were able to match correctly the topic names with its associated lesson color. Test difficulty was less than 20% (p = .1953), and item difficulties for each lesson color ranged from 12% to 30% (see Table 4). A review of the frequency distribution indicates that with several lesson topics, students had misperceptions of what the associated background color was. For example, only 5 participants correctly selected blue as the color associated with the "Need Theories" topic, while 12 selected purple and 13 selected red. These misperceptions were not related to the favorite color: over 53% of the participants selected green as their favorite color. ANOVA revealed that topic-color matching scores were identical for the two recognition groups (means were 1.09). In the recall condition, the mean score was .56 for the context matching group and 1.08 for the context mismatching group. This difference (.53) was not statistically significant, F (1, 42) = 1.59, p = .223.

Table 4

Frequency distribution and item difficulty for topic-color matching test

	Self-efficacy (Orange)	Goal-setting (Red)	Behavioral (Green)	Equity (Purple)	Need (Blue)	Favorite Color
Blue	10	6	17	7	5	6
Green	9	8	8	11	8	23
Orange	13	12	5	8	5	2
Purple	6	8	6	10	12	6
Red	5	9	7	7	13	4
None						2
Item Difficulty	0.3023	0.1860	0.1628	0.2093	0.1163	

n=43, missing data for one participant

Table 5

Topic-Color Matching Scores

	Cueing Strategy Match	Mismatch	Difference	F-Value	P-Value
Recognition					
M	1.09	1.09	0.00	0.00	1.000
SD	1.64	0.83			
n	11	11			
Recall					
M	0.56	1.08	(0.53)	1.59	0.223
SD	1.01	0.59			
n	9	12			

Note: Maximum score = 5
n=43, missing data for one participant

Discussion

This study examined the effects of using color as a contextual retrieval cue for recognition and cued recall task performance. As Figure 1 depicts, the mismatching group outperformed the matching group on both recognition and cued recall task performance. These differences, however, were not statistically significant. Similarly, while males generally outperformed females on both posttests, no statistically significant interactions were found. The data fail to support ICE Theory (Murnane et al., 1999) – insofar as no context effects were observed for recognition tasks – or the outshining hypothesis (Smith, 1986) – recall was not significantly impacted more by context effects than recognition. The general lack of significant differences may be due to the small statistical power of the study, measurement instrument errors, interference from the effects of the practice exercises, or some combination.

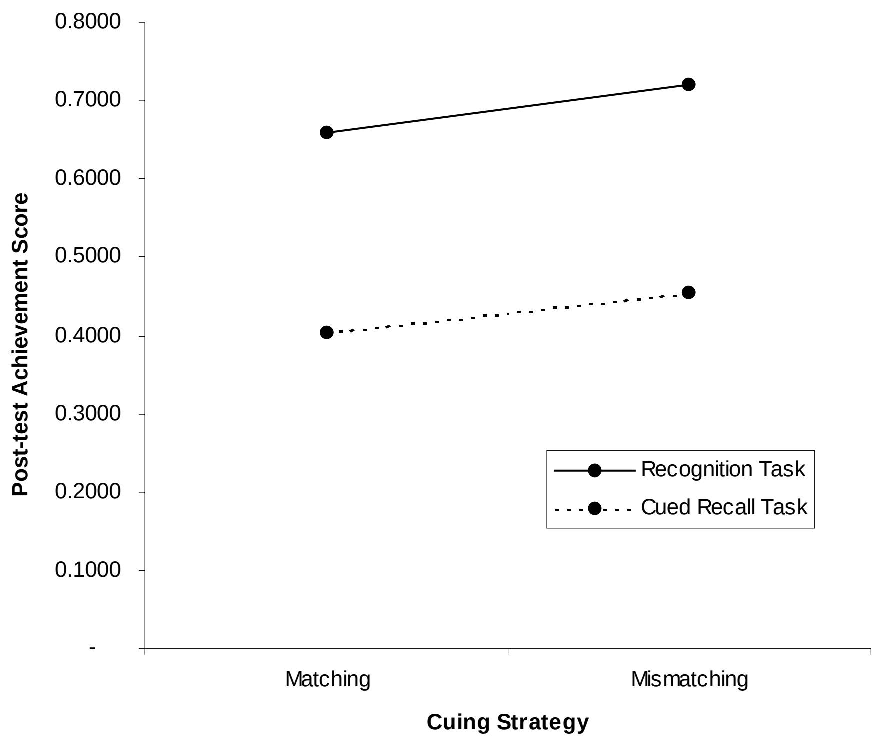

Figure 1. Posttest achievement scores by cueing strategy

Upon examination of the practice scores, it was found that participants in the two recognition groups increased their scores overall by 55%; a 27% gain for the matching group and a 94% gain for the mismatching group. This disproportionate performance improvement for the mismatching group suggests that the practice items, while they did not create a ceiling effect, influenced the posttest scores in a way that favored the mismatching group. Indeed, the mismatch students performed worse on the practice items than the matching group. One possible explanation is that, having scored poorly on the practice items, mismatching students paid closer attention to the feedback given and processed it more deeply than their higher-scoring counterparts in the matching group. This could have accounted for the higher posttest scores.

The recall group did not appear to be affected at all by the practice items during the lesson. Overall gains were zero; however, the posttest mean for the matching group dropped 11% while that of the mismatching group increased 8%. This small difference was enough to swing the posttest results in favor of the mismatching group. Since all of the participants in these two conditions were given the same content and practice items, no within-group differences were expected.

Despite the lack of significant differences, the equivalence of the mean scores point to a surprising possibility: namely, that mismatching context cues could serve the same function as matching context cues in facilitating retrieval functions. In two experiments involving the use of category headings as encoding and retrieval cues, Epstein and Dupree (1977) found that although cued-cued participants (participants were cued with headings both at encoding and at retrieval) scored highest on a cued recall task, uncued-cued participants (participants viewing no headings at encoding but seeing them at retrieval) performed equivalently. They explain this by writing: "… the lack of cueing at recall depressed performance greatly if the items had also not been categorized at input and depressed performance only slightly if the items had been initially categorized" (p.105).

Epstein and Dupree (1977) also found that when context cues were given at encoding but not at retrieval (cued-uncued), participants performed significantly better in free recall tasks than those who were not cued at all (uncued-uncued). They wrote: "Thus, it is seen that in the absence of explicit recall cues, free recall is greatly enhanced by imposed organization during acquisition" (p. 104). As Epstein and Dupree point out, encoding specificity does not account for these organizational effects.

In the present study, all four groups were cued at encoding (input) through the color-coding scheme of the five lessons, the use of headings throughout the lessons, and the use of a main menu that listed all five lessons (categories). Epstein and Dupree's hypothesis, which is based upon a generation-recognition model, would predict equivalent results, regardless of the retrieval cue, since all of the conditions may be influenced by the color-coding's organizational effects. Epstein, Dupree, and Gronikowski's (1979) study; however, failed to support this hypothesis. In this set of experiments, the researchers replicated their earlier study but also manipulated the number of categories (2, 3, 5 and 6 categories). This time, even though cueing improved recall when applied at input, output, and both input and output, they did find significant differences between the cued-cued and the cued-uncued conditions. From these results, they concluded that their study of headings as contextual cues could neither support their earlier generation-recognition hypothesis (since cued-cued did produce significantly greater recall scores than all other cueing strategies), nor could it support a strict interpretation of Tulving's specificity encoding principle (since the other cueing strategies also produced significantly higher recall scores than the non-cued control group).

Implications for Further Research

The results of this study suggest that providing the learner with the opportunity for practice and feedback may inhibit the effects of contextual cues. This possibility requires further study with feedback manipulated as an independent variable. If differences are found, i.e., the with-feedback group experiences lower context effects than the no feedback group, it may also be useful to analyze the results by item difficulty. It is possible that feedback inhibits context effects by increasing the trace strength of the encoded memory, thereby facilitating retrieval with lesser demand for contextual cues, which is somewhat consistent with the outshining hypothesis.

Another possibility suggested by this study is that the contextual effects of color are not strong enough to be pervasive over the course of lengthy instructional content. The instructional unit used in this study employed five lessons and took between 30 and 40 minutes to complete (not including the testing). Over time, as new content and new colors were introduced, the cueing strength of the colors may have dissipated, an alternate explanation for the higher gains in test scores for the mismatching groups. Epstein, Dupree, and Gronikowski (1979) found that the context effects of headings declined as more categories were added. Future studies should account for time as well as the ratio of content-to-number of categories as a way of gauging context cueing strength.

Conclusion

Although the effects of context cues on retrieval functions have been under examination for several decades, few definitive conclusions can be drawn from the research. Both global and local context cues appear to impact recall, although as Epstein and Dupree's studies demonstrate, it is difficult to pinpoint under what conditions context will have the most impact. Also, since any variable that is not a to-be-remembered element of the instruction can be considered a contextual variable, it may not be possible to employ an all-encompassing theory of context. After all, headings, colors, smells, sounds, screen resolution, and even the room in which one experiences the instruction – all potential contextual cues – are very different kinds of factors that may impact different learners in different ways. It is possible that – as Epstein, Dupree, and Gronikowski suggest (1979) imply – the exploration of "situationally specific mini-theories" for contextually cued retrieval is necessary to explain the effects of various types of context cues as they interact with other instructional variables, such as encoding strategies.

Background color, the contextual cue used in this study, is a ubiquitous variable in computer-based instruction, yet there is relatively little research to guide instructional design decisions regarding its usage. Although further researcher is needed, the results of the current study suggest that encoding strategies (such as practice and feedback) moderate the impact of color as a context retrieval cue, particularly with recognition task performance.

Appendix E:

Informed Consent Form

Informed Consent Form

The Pennsylvania State University

Title of Project: Effects of Context on Retrieval Performance

People in Charge:

Dr. Andrew Peck
350 Moore Building
865-1838
acp103@psu.edu

Gustavo Prestera
314 Keller Building
865-1838
gep111@psu.edu

1. This section provides an explanation of the study in which you will be participating:

   a. The study in which you will be participating is part of research intended to assess a) how people learn from multimedia tutorials and b) ways to increase tutorial-based learning. By conducting this study, we hope to improve the use of multimedia for training and instruction.

   b. If you agree to take part in this research, you will be asked to complete some brief questionnaires and take brief comprehension tests. You will view a multimedia presentation. Your answers, along with other participants, will be used to draw conclusions about how technology aids can be used to increase comprehension.

   c. Your participation in this research will take approximately 60-90 minutes.

   d. In return for your participation, you will receive course credit as specified in the syllabus provided by your instructor. Alternative means for earning course credit are available as specified in the syllabus provided by your instructor. You will not be entitled to other compensation.

2. This section describes your rights as a research participant:

   a. You may ask any questions about the research procedures, and these questions will be answered. Further questions should be directed to the people in charge listed above.

   b. Your participation in this research is confidential. Only the people in charge will have access to your identity and to information that can be associated with your identity. In the event of publication of this research, no personally identifying information will be disclosed. To make sure your participation is confidential only a code number appears on your answer sheets and materials. Only the researchers can match names with code numbers.

   c. Your participation is voluntary. You are free to stop participating in the research at any time, or to decline to answer any specific question without penalty.

   d. This study involves minimal risk; that is, no risks to your physical or mental health beyond those encountered in the normal course of everyday life.

3. This section indicates that you are giving your informed consent to participate in the research:

Participant:

I agree to participate in a scientific investigation described above, as an authorized part of the education and research program of the Pennsylvania State University.

I understand the information given to me, and I have received answers to any questions I may have had about the research procedure. I understand and agree to the conditions of this study as described.

To the best of my knowledge and belief, I have no physical or mental illness or difficulties that would increase the risk to me of participating in this study. I am not colorblind.

I understand that I will receive course credit for participating as specified in the syllabus created by my instructor, and that I am entitled to no other compensation.

I understand that my participation in this research is voluntary, and that I may withdraw from this study at any time by notifying the person in charge.

I am 18 years of age or older.

I understand that I will receive a signed copy of this consent form.

_________________________ _______________
Signature Date

Researcher:

I certify that the informed consent procedure has been followed, and that I have answered any questions from the participant above as fully as possible.

_________________________ _______________
Signature Date

Appendix F:

Debriefing Document

Effects of Context on Retrieval Performance

Debriefing

Thank you for participating in this study. The results will be utilized to improve screen design in web-based tutorials. An independent variable is a factor that is systematically manipulated by the researcher to test a research hypothesis. The independent variable in this study was the use of color. Depending on which group you were in, you may have viewed screens that contained a streak of color, which varied in hue depending on the content of a particular screen. That same color was then re-used in the retention test that you took after reading the content. Our hypothesis is that students will recall and recognize content information better if that color cue matches the one used on the screen where they originally learned the material. Other groups did not see color information, or did not see useful color information.

The dependent variable (the factor that we are trying to affect by manipulating the use of color) is test performance. We are looking the participant's ability both to recall content information (through fill-in-the-blank questions) and to recognize information (through multiple choice questions). This is why you were asked to complete a test after each learning unit.

The results of this study will help instructional designers make decisions about how they can use color in their screen design as a strategy for improving learning. Thank you for participating!!! If you have any questions or concerns about this study, please contact:

Dr. Andrew Peck

GUSTAVO E. PRESTERA

Instructional Systems

EDUCATIONAL BACKGROUND

- Ph.D., Instructional Systems with Management minor, The Pennsylvania State University, August, 2003
- M.B.A., The Pennsylvania State University
Graduated: December, 1999
- B.S., Marketing, Saint Joseph's University
Graduated: May, 1992

PROFESSIONAL EXPERIENCE

- Instructional Design Consultant: Independent Contractor, 5/01-Present
- Project Leader/Instructional Designer: *Allen Communication*, 12/99-5/01
- Training and Instructional Design Manager: *Pet Valu International*, 5/99-11/99

TEACHING EXPERIENCE

- Introduction to Web-based Instruction (Spring, 2001). Co-taught at Penn State Great Valley.

PUBLICATIONS

- Moller, L., Prestera, G.E., Harvey, D., Downs-Keller, M., & McCausland, J. (2002). Creating an organic knowledge-building environment within an asynchronous distributed learning context. *The Quarterly Review of Distance Education*, *3*(1).
- Prestera, G.E. & Moller, L. (2001). Organizational alignment supporting distance education in post-secondary institutions. *The Online Journal of Distance Learning Administration* [Online], *4*(4). Available at: http://www.westga.edu/~distance/ojdla/winter44/prestera44.html.
- Prestera, G.E. & Moller, L. (2001). Exploiting opportunities for knowledge building in asynchronous distance learning environments. *The Quarterly Review of Distance Education*, *2*(2), 93-104.