4-26-2017

Examining the Boundaries of the Spacing Effect in Inductive Learning

Chanda Simkin Murphy

EXAMINING THE BOUNDARIES OF THE SPACING EFFECT IN INDUCTIVE LEARNING.

by

Chanda Simkin Murphy

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Psychology

The University of Memphis

May, 2017

Dedication

This degree is dedicated in loving memory to my mother, Mary Claire Simkin. Without her belief and support in me I would have never quit a successful corporate career to pursue my dream of earning a PhD. She was a brilliant woman who put her time and energy into all of her daughters. Even through her final moments of life she was teaching us new things and showing selflessness to everyone around her. I will carry on her legacy through my teaching by continuing to encourage and educate others to be compassionate and selfless. Thank you mom for believing in me for without your encouragement this degree would have never happened.

Acknowledgements

I would like to express my sincerest thanks to Dr. Phil Pavlik for taking me in under his wing to help me finish my PhD. Without his patience and belief in me there would have been no way I would have persisted through finishing this degree. Completing the degree in a different program that was outside of my comfort zone was a major challenge for me but Phil and the rest of the supportive OLL lab made it possible. Thank you to my committee, Dr. Jason Braasch, Dr. James Murphy, and Dr. Rodney Vogl for their guidance, time and encouragement throughout my dissertation process.

Thank you to Kenneth Barideaux and Jackie Maas for their incredible love and support as well. They took me into their lab without question and immediately made me feel like part of the team. Their stories and laughter and patience with all of my incessant questions made the ending of my graduate school experience enjoyable.

Thank you to my Industrial Organizational graduate school crew for getting me through the first half of this degree and continuing to push me to finish the PhD even in the toughest times. An extremely heart felt thank you to my Christian Brothers University family. Not only was the support from my colleagues necessary for this degree but also the support and love from my students was a constant reminder of why I was putting my time and energy into the degree.

Mostly importantly there are never enough words to describe how thankful I am for my family. To my sisters, thank you for picking up the phone every time I needed to vent. To my father, thank you for raising me to be as tough as I am to get through the challenge of this degree. Finally to Joe and Mia, thank you for your patience with me and your undying belief that I could do this. I love you and could never thank you enough.

Abstract

Murphy, Chanda Simkin, Ph.D. The University of Memphis. May, 2017. Examining the Boundaries of the Spacing Effect in Inductive Learning. Major Professor: Philip Pavlik, Jr., Ph.D.

The current study aimed to investigate the role prior knowledge plays in the spacing effect by attempting to replicate the results of two previous studies. Eighty-five participants were divided into two different conditions and practiced diagnosing 36 case studies of six psychological disorders. The only difference between the conditions was whether the participant received the real labels of the disorders (i.e., depression, anxiety, bipolar) or novel labels of the disorders (i.e., wos, baj, pliq). Individual differences in learning strategies were also assessed to examine if there was any relationship between achievement goals, intelligence theories and confidence and the spacing effect. Based on the previous studies, it was hypothesized that there would be an interaction between the spacing effect and label type such that novel labels would produce a stronger spacing effect than known labels. There were no significant differences found for the spacing effect in either the real label or novel label condition leaving the role prior knowledge plays in the spacing effect unconfirmed. The results of the current study necessitate a discussion about the boundaries to the spacing effect and how the most effective use of spaced study can be applied to the classroom.

Table of Contents

List of Tables

Chapter 1

Introduction

Although publications promote best practices for learning and retention, few of these prescribed best practices have been carefully tested outside of a lab setting or with attention to relevant variables. One of these best practices that has been repeatedly studied over the years is spaced study. The spacing effect has been studied with multiple variables ranging from verbatim verbal learning (e.g., Cull, 2000; Janiszewski, Noel, & Sawyer, 2003; Kornmeier, Spitzer, & Sosic-Vasic, 2014; Pavlik & Anderson, 2005) to categorical learning (e.g., Kornell & Bjork, 2008; Wahlheim, Dunlosky, & Jacoby, 2011; Zulkiply, McLean, Burt, & Bath, 2012). One of the major gaps still in this research is on the application of the spacing effect for improving categorical learning in the classroom. One factor that has been purposely left out of this research involving categorical learning and the spacing effect is prior knowledge. Therefore, the purpose of the current study is to further bridge this gap by examining how prior knowledge plays a role in the spacing effect in inductive learning.

Spaced versus Massed practice

Massed study is defined as any study of a topic without interruption or practice of intervening items (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). An often used example of massed study in academia is cramming for a test or, in general, reviewing material with short or no delays between repetitions of the same or similar material. In contrast, spaced study refers to distributed practice in which a measurable amount of time or differing items are interjected between repetitions (Cepeda et al., 2006). An example of spaced practice would be breaking up study over a period of days or weeks leading up

to a test. Wider spacing means having longer delays between repetitions of the same material.

The study of massed versus spaced practice started as early as the 1800's in association with memory and retention research (Ebbinghaus, 1885/1964). Ebbinghaus (1885/1964) found that distributing practice over a span of time provided for better retention in learning a series of syllables. Since then thousands of studies on the spacing effect have been conducted and continue to be conducted in both modern cognitive and educational literatures. These previous studies examined a range of stimuli from verbal memory tasks, such as list recall and paired associates (Cull, 2000; Janiszewski et al., 2003; Kornmeier, et al., 2014; Pavlik & Anderson, 2005), text comprehension (Reder & Anderson, 1982), and categorical assignment of items (Kornell & Bjork, 2008; Wahlheim et al., 2011; Zulkiplyet al., 2012). Previous research also focuses on a number of spacing effect variables including interleaving (e.g., Carvalho & Goldstone, 2012; Wahlheim et al., 2011; Zulkiply & Burt, 2012), embellishment (e.g., Reder & Anderson, 1982), the duration of the spaced interval (e.g., Cull, 2000), age (e.g., Kornell, Castel, Eich, & Bjork, 2010) inductive learning (e.g., Kornell & Bjork, 2008; Zulkiply et al., 2012) and the testing effect (e.g., Cull, 2000; Kornmeier et al., 2014).

Despite all the research that has been done since Ebbinghaus (1885/1964) supporting spaced study over massed study, there is still a disconnect between what is being done in the laboratory and what is being applied in the classroom. In an article by Dempter in 1988, he suggests this failure stems from the lack of alignment between conditions studied in the laboratory and conditions in a classroom. For example, at that time most of the applied studies on the spacing effect focused on simple tasks like text

recall (Dempster, 1986) or vocabulary learning (Dempster, 1987), whereas classrooms usually require more complex learning, and it is not clear whether beneficial effects of spaced study can be extrapolated to complex learning (Dempster, 1988). Similarly, Pashler, Rohrer, Cepeda, and Carpenter (2007) note that many studies have shown benefits of spacing on learning using vocabulary word tests and math problems. However, they were unable to show similar results when examining the spacing effect on inductive learning (i.e., checkerboard patterns, dermatological diagnoses). They also conclude that more parallels are required between laboratory variables and classroom conditions and content. Like Dempster (1988), Rohrer and Pashler (2010) argue that benefits seen using limited study variables, like vocabulary learning (Bahrick, Bahrick, Bahrick, & Bahrick, 1993) and fact or text recall (Carpenter et al. 2009), cannot be generalized to more complex classroom learning. These reviews by Dempster (1988), Pashler et al. (2007) and Rohrer and Pashler (2010) highlight the need to study more complex and applicable stimuli, e.g., categorical assignment or problem solving, in order to establish a better connection between research findings and classroom application.

A study by Kornell and Bjork (2008) was one of the first to test stimuli that better bridged the gap from the lab to the classroom. This paper introduced a new paradigm that showed how spacing affects inductive learning. In contrast to previous research with spacing, they hypothesized that massed practice of category examples is more effective than spaced practice because massed practice allows commonalities to be more easily drawn between features of the categories. Kornell and Bjork's study required the assignment of paintings to the appropriate artist and included both a practice and testing phase. In the practice phase, paintings were randomly assigned to a massed or spaced

presentation, and participants reviewed the painting with the artist's name displayed. In the testing phase, new paintings by the same artists were presented, and participants needed to identify the correct artist's name from multiple choices. With this inductive learning design, Kornell and Bjork discovered, in contrast to their hypothesis, that spaced practice of examples from a category results in better posttest performance than massed practice. Many subsequent categorical learning studies followed Kornell and Bjork's methods but have used different stimuli such as the classification of bird names (Walheim, Dunlosky, & Jacoby, 2011) or butterfly names (Birnbaum, Kornell, Bjork, & Bjork, 2012).

In an effort to support and generalize the findings of Kornell and Bjork (2008) and the other categorical research (Birnbaum et al., 2012; Walheim et al., 2011), Zulkiply et al. (2012) replicated the methods of the aforementioned studies but used case studies of psychological disorders as the categorical stimuli instead of paintings, birds or butterflies. The use of text-based stimuli by Zulkiply et al. seems a notable contribution to the spaced versus massed practice literature due to the educational relevance of learning from text in most academic settings. Zulkiply et al. modeled the practice phase design of Kornell and Bjork (2008) by presenting three case studies for each of six psychological disorders in either spaced or massed presentation. In this practice phase, the participant learned six psychological disorders by reviewing the correct diagnosis and the case study presented on a screen. The test phase then presented unseen case studies and the participant was asked to correctly choose from the same six psychological disorders. To control for prior knowledge, Zulkiply et al. (2012) used novel labels for the disorder names, e.g., Duv was substituted for Obsessive Compulsive Disorder, Tem for

Schizophrenia, Baj for Phobia Disorder, Pliq for Attention Deficit Disorder (Inattentive type), Hix for Attention Deficit Disorder (Hyperactive and Impulsive type) and Wos for Depression. Zulkiply et al. (2012) replicated the findings of Kornell and Bjork (2008) with this new material and similarly conclude that inductive learning benefits from spaced practice.

The Role of Prior Knowledge

By testing college students with stimuli they would normally be learning in a classroom, the Zulkiply et al. (2012) study better bridged the gap between laboratory conclusions and classroom applications. However the use of novel names was unlike what is taught in the classroom and means we cannot be sure the effect would be the same if real names were used. By using these novel names Zulkiply et al. screened out some prior knowledge and created stimuli that were more representative of naïve vocabulary learning. It could be argued that the results of many of the previous studies, which found a significant effect of spaced inductive study over massed study, are confounded by a similar vocabulary learning issue. Instead of finding a spacing effect in relation to inductive learning, these previous studies could actually be finding a spacing effect due to learning unknown labels.

A first example of this issue in a study comes from Kornell and Bjork (2008), who found a spacing effect in their study using classification of artists and their paintings. They found a significant result with spaced study (M = .61) having a higher performance over massed study (M = .35). However, conclusions drawn from their results may have been obscured by their procedure of using poorly known artist names for the study. In this case the spacing effect may have resulted from learning the new names rather than

from learning proper classification of the artist's style. Consider that the chosen artist names were relatively uncommon to those who have not studied art, i.e., Georges Braque, Henri-Edmond Cross, Judy Hawkins, Philip Juras, Ryan Lewis, Marilyn Mylrea, Bruno Pressani, Ron Schlorff, Georges Seurat, Ciprian Stratulat, George Wexler, and Yiemei. It seems plausible that the main performance increase observed in their study was due to spacing effect improving recognition and discrimination (in their multiple choice task) of these previously unfamiliar or unknown names.

A second example is a study by Birnbaum et al. (2012), in which they found a positive spacing effect while testing object recognition and discrimination using butterfly species with names such as Admiral, American, Baltimore, Cooper, Eastern Tiger, Hairstreak, Harvester, Mark, Painted Lady, Pine Elfin, Pipevine, Sprite, Tipper, Tree Satyr, Viceroy, and Wood Nymph. As with Kornell and Bjork, unless the participant had a prior knowledge of butterfly species' names (an amateur lepidopterist), the measured spacing effect could have been due to the learning of the names of the species rather than the perceptual category. This study by Birnbaum et al. (2012) ensured participants had no prior knowledge of the test subjects by changing the names of the butterfly species to one word or if the name described physical characteristics, changing the name entirely. By eliminating the potential for prior knowledge, this study design seems likely to increase the amount of learning needed for word/name acquisition and thus makes the task even more dependent on verbal learning.

Similarly, in another effort to better understand the inductive spacing effect found by Kornell and Bjork (2008), Walheim et al. (2011) studied the learning of bird families. Specifically, Walheim et al. used bird names such as chickadees, finches, flycatchers,

grosbeaks, jays, orioles, sparrows, swallows, thrashers, thrushes, vireos, and warblers. Similar to previously mentioned studies, they found a significant spacing effect. Although some of these names are familiar to many, we think it plausible that many college students have no notion of the difference between a chickadee, a finch and a swallow. Thus the results of this study may also be confounded by lack of prior knowledge, leaving the possibility that learning of the labels was benefitting from spacing effects, and not the learning of categories.

Previous Study

In an effort to replicate the spacing effect produced in the previous studies and bridge the gap from laboratory to classroom, a study was conducted using Zulkiply et al.'s method however we replaced the novel labels for each disorder with the actual names of the disorders (Murphy & Pavlik, accepted). Much like Zulkiply et al. (2012), we conducted a study using applicable categorical stimuli by having the participants study symptoms of psychological disorders and identify the disorders. This previous experiment added an element of testing during the study phase to account for research that has shown testing improves retention (Roediger & Karpicke, 2006; McDaniel Anderson, Derbish, & Morrisette, 2007; McDaniel, Roediger, & McDermott, 2007). In contrast to the findings of Zulkiply et al. (2012), our previous study did not find any significant differences between massed and spaced study. There were also no significant differences found in learning with testing relative to study (this result may be explained by the short retention interval in our experiment, since testing tends mostly to show results after a substantial retention interval). The stimuli that were used in this previous study were analyzed to ensure that properties of the stimuli set were not confounding the

results, for example, the range in the performance on the stimuli of the previous study showed that there was ample room for learning to occur. Finally, the data were also analyzed to examine whether the spacing effect may have had greater impact with either high or low performers by conducting a median split on both posttest scores and prior knowledge scores and there were no significant differences found.

The contrasting results found in the previous study as compared to Zulkiply et al.'s research (2012), are important to the field of learning because it leads us to question the mechanism by which the spacing effects are benefitting learning as reported in prior studies. A difference between Zulkiply et al. and our previous study that could have plausibly led to the differing results is our use of real labels for the disorders as opposed to Zulkiply et al, which used made-up disease labels, such as tem, pliq, and baj. The use of novel labels in the Zulkiply et al. (2012) study produced results like studies on the spacing effect and categorical learning that used novel names such as unknown names of birds, artists and butterflies (Kornell & Bjork, 2008; Walheim et al., 2011; Birnbaum et al., 2012) which also controlled prior knowledge. Taken together, this research begs the question; does prior knowledge (such as knowing disorder labels) negate the effect of spaced study over massed study in inductive learning? If this should be so, there are important implications for how we might use or not use this finding in the classroom.

Learning Process Measures

For the current study, we were interested in getting a better idea of how the students approached performance on a learning task and what sort of difficulties or individual differences might have been related to the spacing effect or overall performance. There has been extensive research on achievement goals and their

relationship to learning outcomes ((Bernacki, Aleven, & Nokes-Malach, 2014; Elliot & McGregor, 2001; Harackiewicz, Barron, & Elliot, 1997). Achievement goals have been shaped and revised over the years but the most recent research has focused on Elliot and McGregor's four constructs (Elliot, 2005). Elliot and McGregor's (2001) achievement goals are comprised of four different achievement goal constructs: performance approach, performance avoidance, mastery approach, and mastery avoidance. Previous research has shown that individuals with performance approach or avoidance goal orientation focus on performance outcomes and social comparison as motivating factors in learning. In contrast those that have a mastery approach or avoidance orientation focus on task mastery and have more of an intrinsic motivation for learning (Elliot & McGregor, 2001). Based on this previous research and the goals of the current study, a survey assessing Elliot and McGregor's (2001) achievement goals will be used to investigate if there is any relationship between the goals and the spacing effect or overall performance.

Another area of interest in the approach to learning that is very often analyzed alongside achievement goals is Dweck's implicit theories of intelligence. Dweck's research has shown that people have two different ways in which they view or understand intelligence. The first theory of intelligence is entity theory in which people view their intelligence as a fixed entity. The other theory of intelligence is incremental theory in which people view their intelligence as malleable (Dweck & Molden, 2000). Research has shown that participants' views on their intelligence can have an effect on their performance based on the task and their individual skill level. Those with an entity theory of intelligence need easy tasks that lead to low effort success so they can appear smart

with no threat to self-esteem, whereas, those with an incremental theory of intelligence need to be challenged and feel like they are putting their knowledge to good use (Dweck & Molden, 2000). To further investigate individual differences on the spacing effect and overall performance; Dweck's assessment on theories of intelligence will also be included in the study (Dweck & Molden, 2000).

Finally, confidence ratings will also be measured in the current study to further investigate how individual differences may be related to the spacing effect or performance. Confidence ratings have been used in many areas of previous research (e.g., Crawford & Stankov, 1996; Stankov, Pallier, Danthiir, & Morony, 2012). In 2012 Stankov et al. found that one's confidence was related to performance and was moderated by ability of the participants and the difficulty of the task. In 2012, Stankov et al. further confirm that confidence is the best predictor of achievement in both math and English. The current study will measure confidence in performance to further examine what relationship individual differences might have with the spacing effect.

Current Study

Based on the previous inductive learning studies using unknown names (Birnbaum et al., 2012; Walheim et al., 2011; Zulkiply et al., 2012) and the contrasting results of the previous study using real disorder names (Murphy & Pavlik, accepted), we theorize that the use of novel or unknown names produces results similar to research on the learning of vocabulary terms, where spacing effects are easy to produce (e.g., Cull, 2000; Janiszewski et al., 2003; Kornmeier et al., 2014; Pavlik & Anderson, 2005). Therefore, we argue the results found in these previous spacing effect studies may be due to the learning of the new terminology and not due to inductive learning. The current

study is designed to provide more evidence as to whether the positive effect of spaced practice is in fact due to label learning or instead categorical learning. The goal of the current study is to replicate and support the findings of Zulkiply et al. (2012) on the effect of spaced presentation when using novel labels as well as replicate our previous study by finding no spacing effect when using real labels for the disorder stimuli. The following is hypothesized:

H1: There will be a strong interaction between the spacing effect and label type such that novel labels will result in more spacing effects than known labels.

H2: In the novel label condition, the spaced condition will perform significantly better on posttest than the massed condition.

Chapter 2

Method

Participants and Design

Eighty-four undergraduates from introductory psychology courses at a small, private university in the mid-south participated voluntarily for extra credit in the course. Fifty-six percent of the participants were female and 45% were male and 100% fell into the age range of 18-25. The majority (66.7%) of the participants were in their freshman year of college with the remaining 16.7% being sophomores, 8% juniors, and 6% seniors.

Replicating and expanding on Zulkiply et al. (2012) and Murphy and Pavlik (accepted), this study is a 2 level between-subjects and 2 level within-subjects design. The between portion of the design included two groups: novel disorder labels versus real disorder labels. The within portion of the design was two levels: spaced versus massed practice. The study protocol included the participants completing a prior knowledge measurement, a study phase, a distracter task, a posttest phase and then final surveys.

The study phase consisted of three case studies for six different psychological disorders (generalized anxiety, depression, obsessive-compulsive, schizophrenia, bipolar, and dissociative identity disorder) totaling 18 case studies. These case studies were randomly assigned by disorder to a massed or a spaced condition for each participant. The order of the study phase conditions was counterbalanced using MSMSMS and SMSMSM (M representing 3 massed trials; S representing 3 spaced trials) to control for ordering effects (see Appendix A for an example of the study phase presentation order).

Each participant was randomly assigned to a novel label condition or a real label condition. Those in the novel label condition received the following novel labels to be used in diagnosis instead of the actual disorder names: Duv, Baj, Tem, Pliq, Hix, and

Wos. Those in the real label condition received the actual disorder names to be used in diagnosis.

The posttest phase included 18 new case studies once again including three case studies per psychological disorder. The case studies were divided among three test blocks with one case study from each disorder presented in each block. The presentation order of the case studies within each block of the posttest phase was randomized for each participant (See Appendix B for an example of posttest presentation order).

Materials

The stimuli for the study included 36 case studies of psychological disorders developed and adapted from different abnormal psychology sources (American Psychiatric Association, 2013; Oltmanns & Emery, 1995). Each case study was between 100 and 120 words in length and included descriptions of symptoms related to each disorder (see Appendix C).

The measures used in this study included three questionnaires: 1) a 30 question prior knowledge assessment which measured the participants' general psychology knowledge (see Appendix D), 2) Elliot and McGregor's 12 item achievement goals assessment (2001) which consisted of 12 questions and measured the participants on four dimensions of achievement goal orientation (see Appendix E), and 3) Dweck's 8 item intelligence theories questionnaire (Dweck & Molden, 2000) measuring their thoughts on entity versus incremental intelligence (see Appendix F). The participants also completed a survey assessing their confidence in diagnosing the disorders (see Appendix G) as well as a few demographic questions (i.e., age, sex, year in school). Those participants in the

novel label condition answered an additional survey to determine if they made any associations between the novel labels and actual disorder labels (see Appendix H).

Procedure

A week prior to the computerized portion of the study, the participants completed the multiple choice, paper and pencil prior knowledge questionnaire. For the remainder of the study, participants were tested in private rooms on computers through the MoFaCTs system (Pavlik, Kelly, & Maass, 2016). In the study phase, participants were presented 18 case studies and were asked to read and study these cases. Each case was presented on the screen with the label of the disorder displayed underneath for a total of 30s. Once the 18 case studies were reviewed, the participants were asked to complete a distracter task in which they answered 15 simple subtraction problems lasting approximately 45s (see Appendix I).

Replicating the original designs of Kornell and Bjork (2008) as well as Zulkiply et al. (2012), the posttest phase began immediately after the distracter task. Participants were shown 18 new case studies they had not already read and were asked to identify the disorder. The participants were presented with one case study at a time on the computer screen and were asked to identify the correct disorder using a set of buttons with either the real disorder names or novel disorder names dependent on condition. Participants received feedback for each response. If the answer was correct, "correct" appeared at the bottom of the screen. If the answer was incorrect, the correct answer was given at the bottom of the screen, and the participant had 10s to review the case study.

After the posttest phase, participants in the novel label condition were asked if they made any associations between the novel labels and real labels. Participants in both

conditions then completed the Achievement Goals and Intelligence Theories questionnaires and were asked how confident they were in their diagnoses of the case studies. Finally the participants filled out a three question demographic survey. All participants were debriefed about the experiment. Because students in the novel label condition were asked to learn novel names for real disorders and because this could potentially impact future learning, as part of the debrief these students were supplied the actual disorder names and were encouraged to use the system with the real disorder names for the rest of the semester to study if they wanted (see Appendix J). Participation in this experiment took approximately 30m.

Chapter 3

Results

A 2 x 2 ANCOVA was conducted to examine if there was an interaction between the novel label and real label condition and the spacing effect with prior knowledge entered as a covariate. There was no significant interaction found, $F(1,72) = 1.31$, $p = .26$. A repeated measures ANOVA was also conducted on the data in both the novel label and the real label conditions to further examine the differences between massed and spaced performance within condition. There were no significant differences in performance between massed and spaced study for the condition using real labels for the disorders, $F(1, 41) = .095$, $p = .76$, (massed study ($M = .79$, 95% CI [.73, .84]), spaced study ($M = .79$, 95% CI [.74, .84])). Also there were no significant differences in performance between massed and spaced study for the condition using novel labels, $F(1, 41) = 1.28$, $p = .27$, (massed study ($M = .34$, 95% CI [.27, .41]), spaced study ($M = .30$, 95% CI [.25, .36])). There was a significant difference in posttest performance between the real label condition ($M = .79$) and the novel label condition ($M = .33$), $t(81) = 13.36$, $p < .001$ with the real label condition scoring higher on average in posttest (see Figure 1).

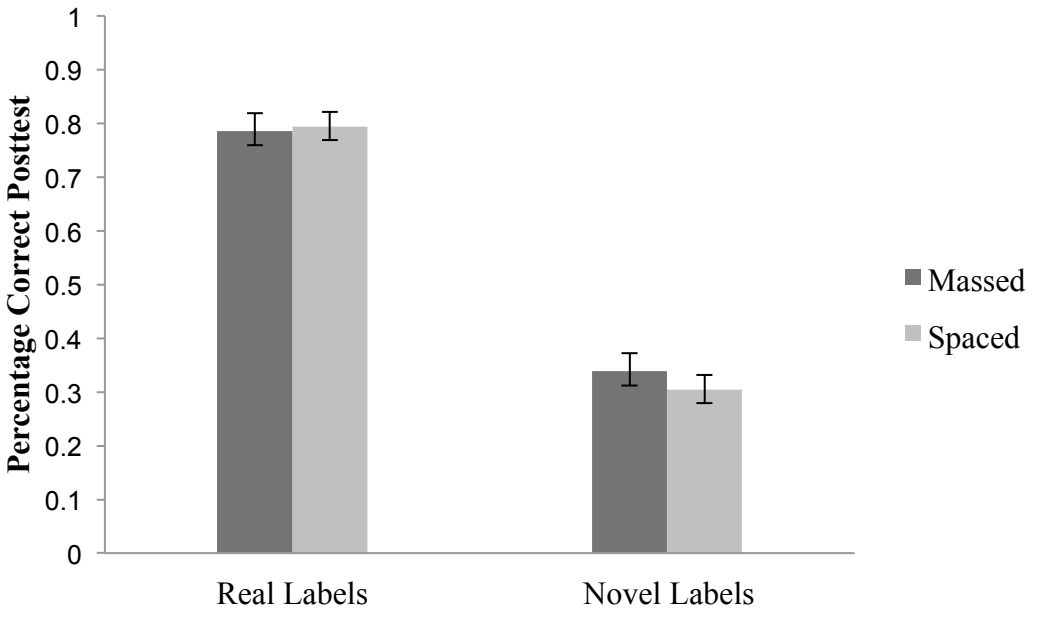

Figure 1. Comparison of probability correct at posttest between massed and spaced performance in both the novel label and real label conditions. Error bars represent standard errors.

A 2 x 2 ANCOVA omnibus test was conducted to see if there was an interaction between group (novel and real) and time spent on the stimuli in massed and spaced conditions. To create a latency variable, average latencies were calculated from the case studies with the correct diagnoses for both the spaced and massed conditions per participant. There was no significant interaction found, $F(1, 68) = .152$, $p = .70$. The data were also analyzed to see if there were any differences in the amount of time participants spent on the stimuli dependent on whether it was a massed or spaced condition within the real and novel label groups. A paired samples t-test was calculated for latencies between the massed and spaced conditions within both the novel and real label conditions. There were no significant differences found in means of massed versus spaced study in either the real, $t(40) = -.51$, $p = .616$, or novel label condition, $t(37) = .28$, $p = .78$. However,

when the average latencies for massed and spaced study were compared between the real and novel label condition, in an independent samples t-test, there were significant differences found in the means. In the real label condition ($M = 27.6$, $SD = 9.4$) participants spent significantly more time on average on those items that were massed than in the novel label condition ($M = 21.2$, $SD = 9.4$), $t(78) = -3.046$, $p = .003$. The same is true for the spaced items, with participants spending more time on average in the real label condition ($M = 28.0$, $SD = 10.3$) than in the novel label condition ($M = 20.7$, $SD = 9.11$), $t(78) = -3.38$, $p = 001$.

The data were also analyzed to look for any trends in misconceptions of diagnoses of the disorders in both the real label condition and the novel label condition. The data reflected the lack of understanding in the novel label condition showing that correct diagnoses of the disorders was not much higher than misconceptions of the disorders (see Table 1 and Figure 2).

Table 1

Percentages of Novel Label Diagnoses

	Participant Answer					
Correct Answer	Baj	Duv	Wos	Hix	Tem	Pliq
Baj	0.25	0.13	0.11	0.17	0.18	0.16
Duv	0.18	0.35	0.13	0.11	0.12	0.12
Wos	0.14	0.11	0.39	0.11	0.14	0.11
Hix	0.14	0.16	0.11	0.31	0.14	0.13
Tem	0.14	0.12	0.08	0.19	0.31	0.17
Pliq	0.15	0.16	0.16	0.12	0.1	0.31

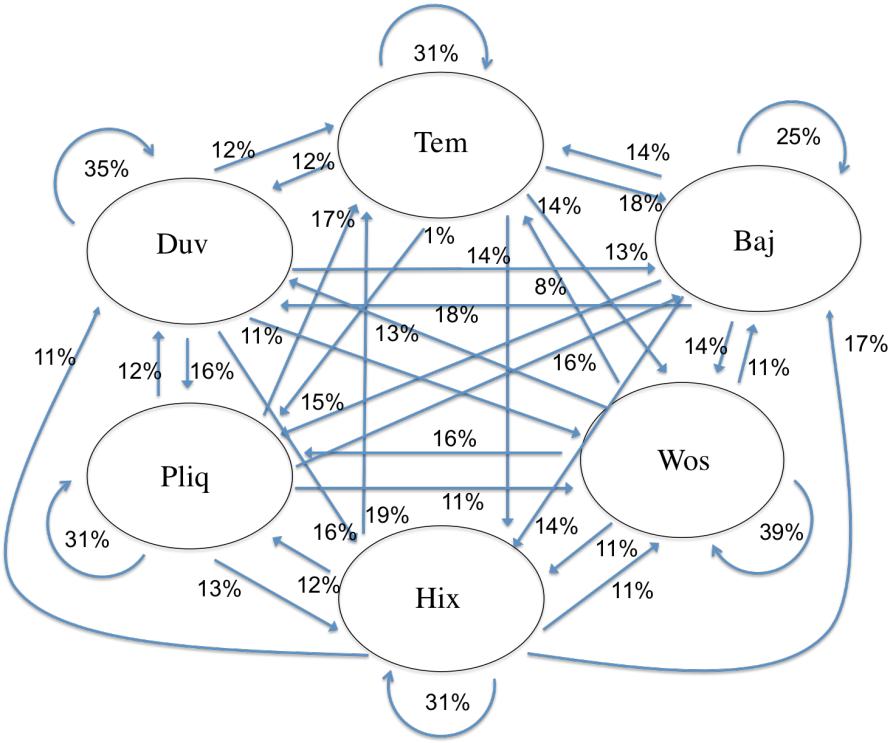

Figure 2. Comparison of percentages of novel label diagnoses.

A chi-square goodness-of-fit test was conducted on each of the disorders in both conditions to see if there were an equal number of errors of each type made for each disorder. For each of the novel label disorders there were no significant chi-square results therefore the misconceptions were equally distributed for each disorder. However in the real label condition, there were significant chi-square results. See Table 2 and Figure 3 for percentages of misconceptions per disorder.

Table 2

Percentages of Real Label Diagnoses

Correct Answer	Participant Answer					
	Anxiety	OCD	Dep	DID	Sch	Bip
Anxiety	0.82	0.09	0.02	0.01	0.04	0.02
OCD	0.05	0.83	0	0	0.05	0.06
Dep	0	0	0.79	0.01	0.06	0.13
DID	0.02	0.02	0.02	0.77	0.06	0.12
Sch	0.02	0.06	0	0.14	0.75	0.04
Bip	0	0.02	0.04	0.09	0.08	0.77

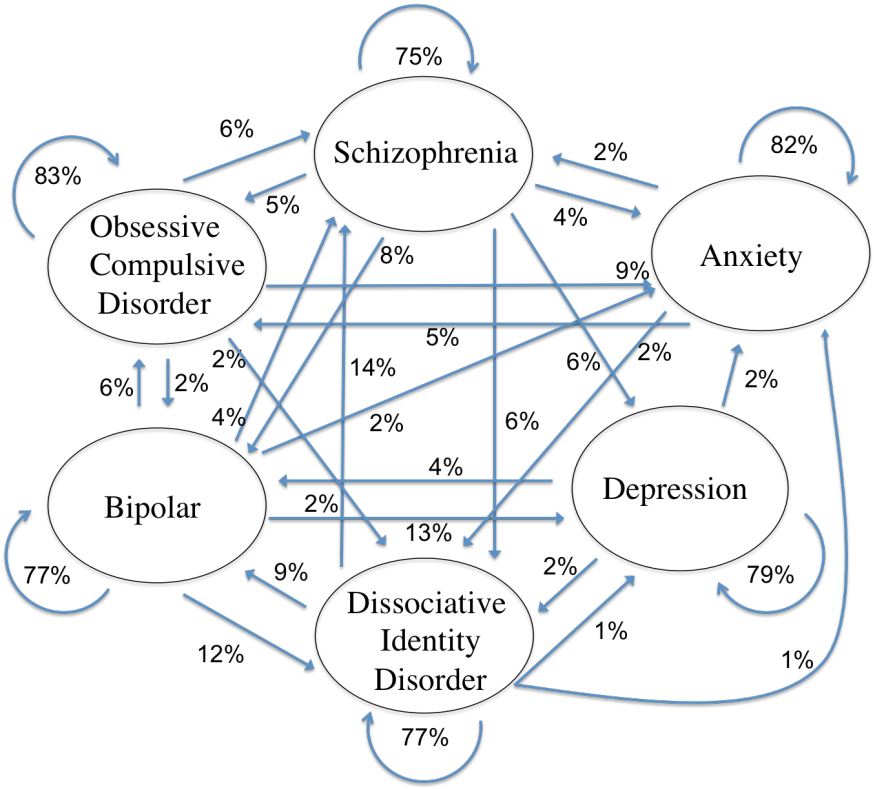

Figure 3. Comparison of percentages of real label diagnoses.

In the real label condition there was an unequal diagnosis of misconceptions for anxiety, X^2 (4) = 13.2, p = .01, depression, X^2 (4) = 44.33, p < .0001, dissociative identity disorder, X^2 (4) = 21.17, p = .00029, and schizophrenia, X^2 (4), = 30.81, p < .0001. A correlation of the matrices was also conducted to see if there was any relationship between the misconceptions of the disorders between the real label and novel label condition. There was no significant relationship between the matrices of misconception proportions between the conditions $r(28) = .09$, $p = .63$.

To further investigate any possible differences in massed versus spaced performance, a spacing effect score was computed for each participant by calculating the difference between the massed and spaced performance on the posttest. Those difference scores were then correlated with scores from multiple variables of interest to the study including prior knowledge, achievement goals, intelligence theories and confidence ratings. There were no significant relationships between the difference scores and scores from any of the aforementioned variables in the real label condition or in the novel label condition. Although there were no significant relationships between the scores on these variables and the spacing effect, average confidence rating scores, achievement goal scores and intelligence theory scores were further examined.

To examine any relationships between prior knowledge and the spacing effect score, an omnibus test of linear regression was conducted to examine the relationship between an aptitude treatment interaction and prior knowledge. Group (novel/real) was entered at step 1 which accounted for .3% of the variance in the spacing effect score however was not significant, $F(1, 73) = .217$, $p = .64$. Prior knowledge was entered at step 2 and explained an additional 6.2% of the variance in the spacing effect score after

controlling for group, *R squared change* = .062, *F change* (1,72) = 4.739, $p = .033$. Although not significant but approaching significance, the total variance explained by the model as a whole was 6.5%, $F(2, 72) = 2.48$, $p = .09$. An independent samples t-test was also conducted to examine if there were any differences between the means of prior knowledge in the real label and novel label conditions. There were no significant differences found between the real label ($M = .53$, $SD = .09$) and novel label ($M = .50$, $SD = .10$) conditions with prior knowledge, $t(73) = 1.13$, $p = .264$.

An independent samples t-test was also conducted on the average confidence rating scores between the novel and real label conditions. There was a significant difference in the confidence rating scores between the novel label ($M = 4.5$) and the real label condition ($M = 2.78$) with the real label condition overall having more confidence in their performance than the novel label condition (with 1 being strongly agree and 6 being strongly disagree), $t(82) = -8.14$, $p < .0001$.

For the current study, a correlational analysis was conducted on the four dimensions of Elliot and McGregor's (2001) achievement goals and posttest scores in both the real label and novel label conditions. See Table 3 for correlational results for the real label condition.

Table 3

Achievement Goals and Posttest Correlations in Real Label Condition

Variables	1	2	3	4	5
1. Perf Approach	—				
2. Perf Avoid	0.056	—			
3. Mastery App	.536**	0.12	—		
4. Master Avoid	.700**	.337*	.654**	—	
5. Posttest	0.073	0.01	0.149	0.124	—
6. Spacing Score	0.196	0.196	0.015	0.032	0.096

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

Achievement goal scores were also correlated with posttest scores in the novel label condition. See Table 4 for the correlational results.

Table 4

Achievement Goals and Posttest Correlations in Novel Label Condition

Variables	1	2	3	4	5
1. Perf Approach	—				
2. Perf Avoid	0.009	—			
3. Mastery App	.513**	.626**	—		
4. Master Avoid	.659**	.357*	.643**	—	
5. Posttest	.372**	0.154	0.345*	0.243	—
6. Spacing Score	0.115	0.071	0.018	0.112	0.076

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

In the novel label condition there was a significant relationship between performance approach scores and posttest scores and mastery approach scores and posttest scores. To further investigate the relationship between achievement goal scores and posttest scores in the novel label condition a repeated measures ANCOVA was used. Performance approach and mastery approach scores were used as a covariate however no significance was found, $F(1, 38) = .376, p = .650$.

To evaluate whether or not intelligence theories had a relationship with participant's performance in the current study, a correlational analysis was conducted. There were no significant relationships between entity theory of intelligence scores, $r(42) = -.164, p > .05, (M = 4.4)$ or incremental theory of intelligence scores, $r(42) = -.031, p > .05, (M = 2.5)$ and posttest in the novel label condition. There were also no significant relationships between entity theory of intelligence scores, $r(42) = .146, p > .05, (M = 4.5)$ or incremental theory of intelligence scores, $r(42) = -.097, p > .05, (M = 2.4)$ and posttest in the real label condition. Additionally there were no significant relationships between entity theory of intelligence scores, $r(42) = .269, p > .05$, or incremental theory of intelligence scores, $r(42) = -.068, p > .05$, and spacing score in the novel label condition. There were also no significant relationships between entity theory of intelligence scores, $r(42) = -.252, p > .05$, or incremental theory of relationship scores, $r(42) = .102, p > .05$, and spacing score in the real label condition A paired sample t-test was conducted to analyze the difference in entity score and incremental scores within the novel label and real label condition. There was a significant difference in the means of intelligence theory scores in the novel label condition with participants having greater incremental scores than entity scores, $t(41) = 7.69, p < .0001$. Like the novel label condition there was a

significant difference in the means of the intelligence theory scores in the real label condition with participants having greater incremental scores than entity scores, $t(41) = 8.84$, $p < .0001$.

In the novel label condition, participants were asked if they made any associations to actual disorders when diagnosing the case studies (e.g., Which disorder below does Duv correspond with?). Frequencies were conducted on how many times a participant made an association between the novel label to a specific disorder, and the results are displayed in Table 5.

Table 5

Percentages of Associations Made

Disorder	Don't know	Anxiety	OCD	Depression	DID	SCH	Bipolar	Other
Baj	0.36	0	0.1	0.1	0.1	0.12	0.14	0.1
Duv	0.26	0	0.33	0.17	0.07	0.07	0.05	0.05
Wos	0.24	0	0.07	0.38	0.02	0.1	0.1	0.1
Hix	0.26	0	0.05	0.05	0.26	0.12	0.19	0.07
Tem	0.21	0	0.07	0.1	0.07	0.29	0.21	0.05
Pliq	0.19	0	0.12	0.36	0.1	0.1	0.14	0

Based on the percentages, the participants made associations to the correct disorder label or didn't make an association at all (i.e., "don't know") more often than making an association to the incorrect disorder. Also important to note that anxiety had zero associations made due to an error in the coding of the computer program, therefore anxiety was not presented as an answer choice for the participants as it should have been.

The associations were analyzed to see if the correct association for each novel label disorder was made above chance. The mean proportions of the incorrect disorder associations were calculated for each disorder. The standard error was then calculated to identify the confidence intervals to establish if the correct association for each disorder was made above chance. For each of the disorders, except Pliq and Baj, a correct association was made above chance. Duv was correctly associated with obsessive-compulsive disorder (M = .33) above chance (M = .089, 95% CI [-.043, .133]). Tem was correctly associated to schizophrenia (M = .29) above chance (M = .113, 95% CI [-.047, .145]). Hix was correctly associated with dissociative identity disorder (M = .26) above chance (M = .10, 95% CI [-.044, .136]). Finally, Wos was correctly associated with depression (M = .38) above chance (M = .04, 95% CI [-.04, .12]). The importance of these results is it shows the participants are making associations to the actual disorders and in most cases the correct disorders.

Chapter 4

Discussion

Textbooks and research papers recommend spaced study as the general best practice for effective studying. However, previous research has failed to take into account several in-classroom variables that make drawing broad conclusions on the effectiveness of spaced study extremely difficult. One major variable that is often left out of spaced study research is the effects of prior knowledge on learning. The current study attempted to directly test whether prior knowledge of material changes the effectiveness of spaced study. Unfortunately based on the results, we still cannot make conclusions about the role prior knowledge plays into the spacing effect. Importantly, the current study was able to replicate the null results of the spacing effect found in Murphy and Pavlik (accepted) however was not able to replicate the spacing effect results found in Zulkiply et al. (2012).

After the completion of the study, an a priori power analysis was computed through G*Power software using the effect size from Zulkiply et al. (2012) and the conservative assumption of no correlation for within-subject values. It was found that a sample of only 18 participants was needed for the spacing effect comparison to achieve .99 power. A sensitivity analysis was also conducted, and it was found that with the sample size of the novel condition of the current study ($N = 42$) we should detect an effect size of $\eta^2 = .06$ with .9 power for the spacing effect comparison. The current study sample size of 42 per condition showed no significant effects of spacing in the novel label condition making conclusions about spacing and the role of prior knowledge

difficult. These results raise important questions about the current recommendations of spaced study as a best practice, which will be specifically addressed.

Misconceptions and Associations

The diagnoses from the posttest were analyzed to investigate any misconceptions that might be occurring. The novel label condition did not have any significant chi square results therefore none of the misdiagnoses of the disorders were made above chance. However, in the real label condition a few of the disorders did have significant chi square results therefore showing the misconceptions made were above chance. For anxiety, participants most often misdiagnosed it as obsessive-compulsive disorder. This misdiagnosis is not surprising as obsessive-compulsive disorder has overlapping symptoms with anxiety disorder and can easily be confused (American Psychiatric Association, 2013). For depression, participants most often diagnosed it as bipolar. Since bipolar has a depression component it is not unexpected that case studies could have been confused as bipolar, especially if not read carefully (American Psychiatric Association, 2013). Dissociative identity disorder was most often misdiagnosed as bipolar. In the previous study (Murphy & Pavlik, accepted), dissociative identity disorder was most often misconceived as schizophrenia, which seems to be a more common misconception than bipolar. However, the misconception in this study could be due to the general complicated nature of dissociative identity disorder (American Psychiatric Association, 2013). Finally, although dissociative identity disorder was not misconceived as schizophrenia, schizophrenia was most often misdiagnosed as dissociative identity disorder. The misconceptions in the current data seem to align themselves with what

would be most commonly confused due to the true overlap of symptoms in these disorders.

An original goal of the current study was to also examine if the participants in the novel label condition were processing an extra step in their decision making by checking to see if they made any associations between the novel label and the real labels. The current results confirmed that participants were not only making associations to actual disorders, but also were making associations to the correct disorders. This is important because Zulkiply et al. (2012) claimed that they were screening out prior knowledge in their study. Based on the associations made in the current study this claim might not be the case. If the participants are making associations to the actual disorders then prior knowledge is still playing a role in their learning. These associations need to be investigated further in future research as a possible prior knowledge variable that may create a boundary to the spacing effect.

Learning Goals and the Spacing Effect

Interestingly the only significant results found between conditions fell into the novel label condition. Data from Elliot and McGregor's (2001) achievement goal surveys conducted in this study enabled us to measure the association of achievement goals and performance. There was a significant positive correlation found between the performance approach scale and posttest as well as mastery approach scale and posttest in the novel label condition. Bernacki et al. (2014) discuss in their research that achievement goals will vary based on not only personal interests of material being learned but also the instructions or goals set up for the task being completed. They state that when the topic is interesting to the person then mastery approach would be more likely to be a goal of the

individual. However, if the instructions given for a task include that their performance will be evaluated to assess competency, then performance approach would most likely be the goal. Harackiewicz et al. (1997) also studied how situational factors can affect goals that students adopt in classrooms. They discuss the adoption of achievement goals in the college setting may be different based on the course level, course content or the way the content is delivered. For example, a performance goal outcome might be more beneficial in a specific situation such as a class that is introductory and the main concern to the students is their grades to get into future classes. Harackiewicz et al. studied introductory psychology courses over the duration of a semester and measured the variables of achievement goals, competence and interest to better understand how individual differences and context play a role in achievement goals. Their results showed that performance goals had a significant positive relationship with final grades, supporting their idea that performance goals are more likely to be adopted in a situational context in which the final grade is the main focus. In the current study, performance approach had a significant positive correlation with posttest scores in the novel label condition. Since the task in the current study was specific to psychology and given to students in an introductory psychology class, it is not surprising based on Harackiewicz et al. results that the there was a significant relationship between performance approach and posttest. It is surprising that mastery approach was significant with posttest due to the labels being meaningless in the novel label condition. However based on Bernacki et al. (2014) an argument as to why mastery approach did have a significant relationship with posttest is because the students possibly found the topic interesting. Psychological disorders are something that is typically intriguing to students, therefore they may have made more of

an effort to master the material because of their interest in it. It is also surprising that there was not a significant relationship between performance approach and mastery approach and posttest in the real label condition as well. A possible explanation as to why there were no significant relationships between achievement goals and posttest in the real label condition is the possibility of a ceiling effect in the posttest scores. With the mean performance at approximately 80% in the real label condition, there was less room for discrimination among participant's scores.

Dweck's research has also shown that people have two different ways in which they view or understand intelligence (Dweck & Molden, 2000). The first theory of intelligence is entity theory in which people view their intelligence as a fixed entity. The other theory of intelligence is incremental theory in which people view their intelligence as malleable (Dweck & Molden, 2000). Research has shown that participants' views on their intelligence can have an effect on their performance based on the task and their individual skill level. Those with an entity theory of intelligence need easy tasks that lead to low effort success so they can appear smart with no threat to self-esteem whereas those with an incremental theory of intelligence need to be challenged and feel like they are putting their knowledge to good use (Dweck & Molden, 2000). Previous research also integrates the study of these implicit intelligence theories with achievement goals finding mixed results in the relationships between achievement goals and entity and incremental theories (Blackwell, Trzesniewski, & Dweck, 2007; Dupeyrat & Marine, 2004; Hong, Chiu, Dweck, Lin, & Wan, 1999). This previous research also had mixed results, with some finding significance and others not, in their findings examining the effect of entity and incremental theories on outcomes and performance. Due to the inconsistencies in the

previous research, these results may support the lack of significant relationships in intelligence theories, achievement goals, and posttest in the current study. Also there is little research that has looked at intelligence theories on this brief of a task. Although the participants were asked questions regarding their overall beliefs of intelligence, they may not have been able to answer those questions beyond the task at hand considering the survey was completed at the end of the study.

Another result only found in the novel label condition was a significant positive relationship between the confidence ratings and posttest. Once again a possible reason for seeing significant results in the novel label condition and not in the real label condition is the mean average for the novel label condition at posttest was approximately thirty-five percent which is much further from ceiling than the real label condition. However another possible cause for this finding is that confidence ratings were collected after the testing portion of the experiment in which participants had been given immediate feedback on their results. Therefore the confidence ratings could reflect that participants in the novel label conditions knew they performed poorly and had a more realistic assessment of their performance rather than the typical overconfidence bias.

The Boundaries of the Spacing Effect

An argument as to why there was no significant spacing effect found in either our novel naming or actual naming groups stems from research regarding the difference between temporal spacing and interleaving. Some previous research has used the terms interleaving and spacing interchangeably due to the explanation that if topics are interleaved (or interchanged) while studying then by definition they are temporally spaced as well. However, some previous research suggests in inductive learning the

spacing effect could actually be attributed to the interleaving of topics and not time between topics. The important difference between temporal spacing and interleaving is the argument that interleaving allows for discriminability of concepts or categories when items are interchanged because an individual can compare the difference between items or categories (Carvalho & Goldstone, 2012; Kang & Pashler, 2012; Mitchell, Nash, & Hall, 2008; Taylor & Rohrer, 2010). In 2010, Taylor and Rohrer did a study to further investigate the interleaving effect as compared to the spacing effect. They had children study math formulas and they had to choose the correct formula to solve the missing value in a shape presented on the screen. In an interleaved condition, the different types of math problems alternated presentations with very little time between presentations. In the massed condition all of the same problems were presented back to back consistent with previous research. Taylor and Rohrer found that the interleaved condition produced better posttest scores. Most importantly the study showed the importance of interleaving when discriminability of categories is a concern. Taylor and Rohrer also argue that if categories are easily distinguished from one another interleaving might be less beneficial.

Kang and Pashler (2012) also argue that the spacing effect could also be attributed to more of an interleaving effect than temporal spacing effect due to the need to discriminate or contrast between categories in inductive learning. In their study they used the assignment of painters to their paintings much like Kornell and Bjork (2008) but used paintings from only three artists. Kang and Pashler investigated this difference in interleaving versus temporal spacing by creating four conditions: a massed condition identical to previous research, a simultaneous same condition in which four paintings by the same artist were given at one time, an interleaving condition in which the

presentations of the paintings alternated between the three artists, and a temporal spacing condition in which the presentation of the paintings from each painter were spaced apart with the material between presentations being "filler" material such as cartoons or a blank screen. Kang and Pashler found significantly higher performance in the interleaved condition than the other three conditions. A second study they did further argues the importance of discriminability between categories by replicating their previous experiment except instead of simultaneous same; they had a simultaneous difference condition. In this condition three paintings were displayed on the screen by the three different artists. In this experiment the simultaneous difference conditions produced the highest posttest results, closely followed by the interleaving condition. Kang and Pashler argue that since the simultaneous presentation of different artists provide as much benefit in learning as interleaving, then the opportunity to discriminate or contrast between categories is important to improving induction.

To further examine how discriminability of categories plays a role in the spacing effect, Carvalho and Goldstone's (2012) built on the previous research of Taylor and Rohrer (2010) and Pashler (2012) by looking at differing complexity of stimuli and the interleaving effect. In their experiments they compared stimuli (different abstract drawings) that had high within category similarity and low within category similarity. Their findings explained how interleaved study could improve learning stimuli with high similarity because smaller differences that are difficult to detect can be more easily distinguished with spacing. However items that have low similarity benefit from massed practice because commonalities can be recognized more easily when stimuli are seen back to back. This supports the argument by Taylor and Rohrer (2010) that the

interleaving benefit may lesson if the categories are highly discriminable (i.e., low similarity). It can be argued the stimuli in the current study fall in the middle of high and low similarity. The case studies had high similarity in their overall symptoms of each within category disorder. However, each case study also had enough variation in details of those symptoms that within category disorders could also be considered low similarity. The fact that stimuli in the current study were neither especially high nor low similarity could contribute to why no difference was observed in massed or spaced study.

The results of this study further support the need for more research surrounding the boundary conditions of the spacing effect. The spacing effect has been considered a best practice for some time, is written into textbooks, and has been recommended to educators in numerous publications. However, the results of the current study and aforementioned previous research support the idea that the spacing effect is highly affected by many different variables such as discriminability of stimuli (e.g., Carvalho & Goldstone, 2012; Kang & Pashler, 2012; Taylor & Rohrer, 2010) and the types of stimuli used (i.e., paintings versus case study analysis) (e.g., Kornell & Bjork, 2008; Zulkiply et al. 2012). Pashler et al. (2007) first discussed these boundary conditions for the spacing effect when they conducted multiple studies with varying stimuli. They were able to find spacing to be significantly more beneficial than massed learning in multiple experiments that used different stimuli including the learning of vocabulary, the learning of unknown facts, and the learning of math facts. However, when they tried the same experiments using the stimuli of checkerboard patterns and dermatological diagnoses, they were not able to find a significant spacing effect. Although many researchers have since been able to find the spacing effect with perceptual criteria similar to checkerboard patterns or

dermatological diagnoses (i.e., paintings, Kornell & Bjork, 2008; butterflies, Birnbaum et al., 2012; and birds, Walheim et al., 2011) based on the results of the current study, we conclude that Pashler et al.'s argument of the existence of boundary conditions to the spacing effect is still valid. Further research is needed on these boundaries to establish when the spacing effect is truly effective in an educational setting and with specifically what types of educational topics. As Dempster originally discusses in 1988, there still seems to be too much of a gap between the laboratory and the classroom especially where the variable of prior knowledge is concerned. Since previous studies on the spacing effect, from vocabulary learning to the diagnoses of disorders, have all screened out prior knowledge, we lack the ability to discern if prior knowledge blocks the spacing effect or not. Additionally, the results of the current study and the previous research discussed call for more research to determine the relationship between stimuli discriminability and the spacing effect. Along with discriminability more research is needed to understand what is actually creating the spacing effect (i.e., temporal spacing or interleaving) in some inductive learning and not others. Another area that needs further research in relation to the spacing effect is the retention interval included before recall of the information studied. In previous studies with verbal learning, the spacing effect has been tested at both short term and long term (i.e., at least 24hr intervals (e.g., Bahrick et al., 1993; Karpicke & Roediger III, 2007) finding that the spacing effect showed for better long-term retention. However in the studies that have looked at the spacing effect in inductive learning there has not been testing at a long-term retention interval. Therefore future research involving the spacing effect and inductive learning should include a long term retention interval to see if the spacing effect would surface in long term retention even if

there is no benefit in the short term. This study successfully replicates the null results of Murphy and Pavlik (accepted) and fails to replicate Zulkiply et al. (2012), thus this study parallels other research that has failed to find the spacing effect. Taken together, these results argue against the continued advocacy of spacing as a default best practice for studying all types of educational material and argues for the importance of continuing to research the boundaries to the spacing effect.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*(5), 316-321.

Bernacki, M. L., Aleven, V., & Nokes-Malach, T. J. (2014). Stability and change in adolescents' task-specific achievement goals and implications for learning mathematics with intelligent tutors. *Computers in Human Behavior*, *37*, 73-80.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392-402.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246-263.

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U. S. history facts. *Applied Cognitive Psychology*, *23*, 760–771.

Carvalho, P. F., & Goldstone, R. L. (2012). Category structure modulates interleaving and blocking advantage in inductive category acquisition. In *Proceedings of the thirty-fourth annual conference of the cognitive science society,*186-191.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215-235.0

Dempster, F. N. (1986). Spacing effects in text recall: An extrapolation from the laboratory to the classroom. *Journal of Educational Psychology*, *79*, 162-170.

Dempster, F N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162-170.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results o psychological research. *American Psychologist*, *43*(8), 627-634.

Dupeyrat, C., & Marine, C. (2005). Implicit theories of intelligence, goal orientations, cognitive engagement, and achievement: A test of Dweck's model with returning to school adults. *Contemporary Educational Psychology, 30*, 43-59.

Dweck, C., & Molden, D. C. (2000). Self theories. *Handbook of Competence and Motivation*, 122-140.

Ebbinghaus, H. (1885/1964). Memory: A contribution to experimental psychology (H. A. Ruger, C. E. Bussenius, & E.R. Hilgard, Trans.). New York, NY: Dover Publications, Inc.

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. *Handbook of Competence and Motivation*, *16*(2005), 52-72.

Elliot, A. J., & McGregor, H. A. (2001). A 2× 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501-519.

Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology*, *73*(6), 1284-1295.

Hong, Y. Y., Chiu, C. Y., Dweck, C. S., Lin, D. M. S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology*, *77*(3), 588-599.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A Meta-analysis of the Spacing Effect in Verbal Learning: Implications for Research on Advertising Repetition and Consumer Memory. *Journal of Consumer Research, 30(1),* 138-149.

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*(1), 97-103.

Karpicke, J. D., & Roediger III, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 704-719.

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories Is Spacing the "Enemy of Induction"? *Psychological Science*, *19*(6), 585-592.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*(2), 498-503.

Kornmeier, J., Spitzer, M., & Sosic-Vasic, Z. (2014). Very Similar Spacing-Effect Patterns in Very Different Learning/Practice Domains. *PloS one*, *9*(3), e90656.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4-5), 494-513.

McDaniel, M. A., Roediger, H. L., & Mcdermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*(2), 200-206.

Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 237-242.

Murphy, C. S., & Pavlik, P. I. (2016). *Effects of Spacing and Testing on Inductive Learning.* Manuscript accepted for publication.

Oltmanns, T. F., & Emery, R. E. (1995). *Abnormal Psychology*. Prentice-Hall, Inc.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*, *14*(2), 187-193.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation Based Model of the Spacing Effect. *Cognitive Science*, *29*(4), 559-586.

Pavlik Jr., P. I., Kelly, C., & Maass, J. K. (2016). Using the mobile fact and concept training system (MoFaCTS). In A. Micarelli & J. Stamper (Eds.), *Proceedings of the 13th International Conference on Intelligent Tutoring Systems* (pp. 247-253). Switzerland: Springer.

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition*, *10*(2), 97-102.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, *39*(5), 406-412.

Stankov, L., & Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personality and Individual Differences*, 21(6), 971–986.

Stankov, L., Pallier, G., Danthiir, V., & Morony, S. (2012). Perceptual underconfidence: A conceptual illusion? *European Journal of Psychological Assessment*, 1–11.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*(6), 837-848.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, *39*(5), 750-763.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*(1), 16-27.

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction, 22,* 215–221.

Appendix A

Examples of Study Phase Conditions

Real label condition:

Condition 1 (MSMSMS)	Condition 2 (SMSMSM)
Depression	Bipolar
Depression	Obsessive Compulsive Disorder
Depression	Depression
Anxiety	Anxiety
Schizophrenia	Anxiety
Dissociative Identity Disorder	Anxiety
Obsessive Compulsive Disorder	Bipolar
Obsessive Compulsive Disorder	Obsessive Compulsive Disorder
Obsessive Compulsive Disorder	Depression
Anxiety	Schizophrenia
Schizophrenia	Schizophrenia
Dissociative Identity Disorder	Schizophrenia
Bipolar	Bipolar
Bipolar	Obsessive Compulsive Disorder
Bipolar	Depression
Anxiety	Dissociative Identity Disorder
Schizophrenia	Dissociative Identity Disorder
Dissociative Identity Disorder	Dissociative Identity Disorder

M = Massed study
S = Spaced study

Novel label condition:

Condition 1 (MSMSMS)

Baj
Baj
Baj
Duv
Wos
Tem
Pliq
Pliq
Pliq
Duv
Wos
Tem
Hix
Hix
Hix
Duv
Wos
Tem

Condition 2 (SMSMSM)

Hix
Pliq
Baj
Duv
Duv
Duv
Hix
Pliq
Baj
Wos
Wos
Wos
Hix
Pliq
Baj
Tem
Tem
Tem

M = Massed study
S = Spaced study

Appendix B

Example of Posttest Blocks

Real Label Condition:

Test block 1

Anxiety
Depression
Dissociative Identity Disorder
Obsessive Compulsive Disorder
Bipolar
Schizophrenia

Test block 2

Obsessive Compulsive Disorder
Schizophrenia
Depression
Bipolar
Dissociative Identity Disorder
Anxiety

Test Block 3

Bipolar
Obsessive Compulsive Disorder
Schizophrenia
Anxiety
Dissociative Identity Disorder
Depression

Novel Label Condition:

Test block 1

Duv
Baj
Tem
Pliq
Hix
Wos

Test block 2

Pliq
Wos
Baj
Hix
Tem
Duv

Test Block 3

Hix
Pliq
Wos
Duv
Tem
Baj

Appendix C

Sample Case Study

Real Labels Condition:

Karen Rusa, 30 years old, is a married woman and a mother of four children. For the past several months Karen has been experiencing intrusive, repetitive thoughts that center around her children's safety. Karen also has noted that her daily routine is seriously hampered by an extensive series of counting rituals that she performs throughout each day. She has described herself as tense, jumpy and unable to relax. She has also reported dissatisfaction with her marriage and problems in managing her children. During the past several weeks, she has been spending more and more time crying and hiding alone in her bedroom (Oltmanns et al. 1991).

Psychological disorder type: Obsessive Compulsive disorder

Novel Labels Condition

Karen Rusa, 30 years old, is a married woman and a mother of four children. For the past several months Karen has been experiencing intrusive, repetitive thoughts that center around her children's safety. Karen also has noted that her daily routine is seriously hampered by an extensive series of counting rituals that she performs throughout each day. She has described herself as tense, jumpy and unable to relax. She has also reported dissatisfaction with her marriage and problems in managing her children. During the past several weeks, she has been spending more and more time crying and hiding alone in her bedroom (Oltmanns et al. 1991).

Psychological disorder type: Pliq

Appendix D

Prior Knowledge Assessment

1) Which branch of psychology is most directly concerned with the study of how people think about, influence, and relate to one another?
   a) developmental psychology
   b) social psychology
   c) personality psychology
   d) clinical psychology

2) Pets who learn that the sound of an electric can opener signals the arrival of their food illustrate
   a) shaping.
   b) extrinsic motivation.
   c) classical conditioning.
   d) observational learning.

3) Jabar, a 25-year-old auto mechanic, thinks he is Napoleon. He further believes he is being imprisoned against his will in the psychiatric hospital where his relatives have brought him for treatment. Jabar is most likely suffering from
   a) obsessive-compulsive disorder
   b) schizophrenia
   c) panic disorder
   d) dissociative identity disorder

4) A generalized anxiety disorder is characterized by
   a) offensive and unwanted thoughts that persistently preoccupy a person.
   b) a continuous state of tension, apprehension, and autonomic nervous system arousal.
   c) hyperactive, wildly optimistic states of emotion.
   d) alternations between extreme hopelessness and unrealistic optimism.

5) Mary enjoys socializing with friends and talking with them on her cell phone. Eileen prefers quiet times by herself when she can reflect on her own thoughts. The characteristics of Mary and Eileen indicate that each has a distinctive
   a) fixation
   b) personality
   c) reaction formation
   d) collective unconscious

6) Which therapeutic approach emphasizes that people are often disturbed because of their negative interpretations of events?
   a) client-centered therapy
   b) systematic desensitization
   c) cognitive therapy
   d) light exposure therapy

7) A mental set is most likely to inhibit
   a) confirmation bias.
   b) overconfidence.
   c) creativity.
   d) belief perseverance.

8) Participants in the Milgram obedience studies were ordered to
   a) play the role of the prison guards.
   b) write an essay supporting a position they didn't believe in.
   c) deliver electric shocks to a learner for giving incorrect answers.
   d) participate in a team tug-of-war by pulling on a rope as hard as they could.

9) Sluggishness and inactivity are most likely to be associated with
   a) antisocial personality disorder
   b) major depressive disorder
   c) obsessive-compulsive disorder
   d) dissociative identity disorder

10) Who is the best example of a Type A personality?
   a) A) Valentin, a self-confident, intelligent journalist
   b) B) Kane, a relaxed, easygoing mail carrier
   c) C) Philip, a competitive, hot-tempered corporation president
   d) D) Thomas, an introverted, inhibited mental patient

11) Kentaro hates to wear ties but wears one to his sister's wedding to avoid his family's disapproval. Kentaro's behavior exemplifies the importance of
   a) the mere exposure effect.
   b) informational social influence.
   c) normative social influence.
   d) social facilitation.

12) Behaving with unselfish concern for the welfare of others is called
   a) social facilitation.
   b) passionate love.
   c) groupthink.
   d) altruism.

13) A chess-playing computer program that routinely calculates all possible outcomes of all possible game moves best illustrates problem solving by means of
a) the availability heuristic.
b) belief perseverance.
c) an algorithm.
d) framing.

14) Systematic desensitization involves
a) depriving a client access to an addictive drug
b) associating unwanted behaviors with unpleasant experiences
c) replacing a positive response to a harmful stimulus with a negative response
d) associating a pleasant relaxed state with anxiety-arousing stimuli

15) Coping refers to a variety of methods used to
a) avoid the adaptation-level phenomenon.
b) inhibit the fight-or-flight reaction.
c) prevent the release of lymphocytes.
d) alleviate stress.

16) Cecil is preoccupied with thoughts of jumping out the window his tenth-floor apartment. To reduce his anxiety, he frequently counts his heartbeats aloud. Cecil would most likely be diagnosed as experiencing
a) panic disorder
b) bipolar disorder
c) generalized anxiety disorder
d) obsessive-compulsive disorder

17) After experiencing inescapable brutalities as a prisoner in a Nazi concentration camp, Mr. Sternberg became apathetic, stopped eating, and gave up all efforts to physically survive the ordeal. Mr. Sternberg's reaction most clearly illustrates
a) a Type A personality.
b) the adaptation-level phenomenon.
c) learned helplessness.
d) an internal locus of control.

18) Those with the narcissistic personality disorder are likely to be preoccupied with
a) an irrational fear of people
b) delusions of persecution
c) physical symptoms of distress
d) their own self-importance

19) The recall of sad experiences is often primed by feelings of sadness. This most clearly illustrates
a) the serial position effect.
b) retroactive interference.
c) the misinformation effect.
d) mood-congruent memory.

20) Abraham Maslow suggested that those who fulfill their potential have satisfied the need for
a) reciprocal determinism
b) self-actualization
c) immediate gratification
d) unconditional positive regard

21) When an individual is unaware that they present different personalities to the world this is knows as
a) schizophrenia
b) dissociative identity disorder
c) antisocial personality disorder
d) narcissistic personality disorder

22) According to Freud, defense mechanisms are used by the
a) id to defend against the accusations and guilt feelings produced by the superego.
b) ego to prevent threatening impulses from being consciously recognized.
c) superego to prevent expression of sexual and aggressive drives.
d) id, ego, and superego in a repetitive sequence of internal conflicts.

23) A person who can imagine many alternative uses of a paper clip best illustrates
a) fluid intelligence.
b) divergent thinking.
c) crystallized intelligence.
d) convergent thinking.

24) Alex experiences little stress because he expects things to work out the way he wants them to. This best illustrates the value of
a) a Type A personality.
b) an external locus of control.
c) optimism.
d) the general adaptation syndrome.

25) George Frideric Handel composed his *Messiah* during three weeks of intense, creative energy. Many believe Handel suffered a mild form of
a) agoraphobia
b) a dissociative disorder
c) bipolar disorder
d) catatonia

26) According to Freud, the part of personality that represents our sense of right and wrong and our ideal standards is the
   a) Oedipus complex.
   b) ego.
   c) id.
   d) superego.

27) The cocktail party effect provides an example of
   a) neuroadaptation.
   b) REM rebound.
   c) selective attention.
   d) hypnagogic sensations.

28) Freud's theory of personality has been criticized because it
   a) underestimates the importance of biological contributions to personality development.
   b) is contradicted by recent research demonstrating the human capacity for destructive behavior.
   c) is overly reliant upon observations derived from Freud's use of projective tests.
   d) offers few testable hypotheses that allow one to determine its validity.

29) One good alternative to antidepressant drugs is
   a) aerobic exercise.
   b) psychosurgery.
   c) virtual reality exposure therapy.
   d) EMDR.

30) Chunking refers to
   a) getting information into memory through the use of visual imagery.
   b) the organization of information into meaningful units.
   c) the unconscious encoding of incidental information.
   d) the tendency to recall best the first item in a list.

Appendix E

Achievement Goals Assessment

Using the scale below, please select the extent to which you agree or disagree with each of the following statements regarding your academic performance by choosing the corresponding button below.

1	2	3	4	5	6	7
Strongly Agree	Agree	Mostly Agree	Neither Agree or Disagree	Mostly Disagree	Disagree	Strongly Disagree

_____ 1. My fear of performing poorly is often what motivates me.

_____ 2. Sometimes I am afraid that I will not understand the content of a class as thoroughly as I'd like.

_____ 3. I am often concerned that I will not learn all that there is to learn.

_____ 4. My goal is to avoid performing poorly.

_____ 5. I want to learn as much as possible.

_____ 6. I just want to avoid doing poorly.

_____ 7. It is important for me to do well compared to others.

_____ 8. My goal is to get a higher score than most of the other students.

_____ 9. It is important for me to do better than other students.

_____ 10. I desire to completely master material presented.

_____ 11. I worry that I will not learn all that I possibly can.

_____ 12. It is important for me to understand content as thoroughly as possible.

Appendix F

Intelligence Theories Assessment

Using the scale below, please indicate the extent to which you agree or disagree with each of the following statements by clicking the button below that corresponds to your opinion.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

_____ 1. You have a certain amount of intelligence, and you can't really do much to change it.

_____ 2. Your intelligence is something about you that you can't change very much.

_____ 3. No matter who you are, you can significantly change your intelligence level.

_____ 4. To be honest, you can't really change how intelligent you are.

_____ 5. You can always substantially change how intelligent you are.

_____ 6. You can learn new things, but you can't really change your basic intelligence.

_____ 7. No matter how much intelligence you have, you can always change it quite a bit.

_____ 8. You can change even your basic intelligence level considerably.

Appendix G

Confidence Ratings Assessment

Using the scale below, please select the extent to which you agree or disagree with each of the following statements regarding your performance during this experiment by choosing the corresponding button below.

1	2	3	4	5	6
Strongly Agree	Agree	Mostly Agree	Mostly Disagree	Disagree	Strongly Disagree

Novel Label Condition

1. I was confident in this assignment when I diagnosed a case study with the disorder of Duv.
2. I was confident in this assignment when I diagnosed a case study with the disorder of Baj.
3. I was confident in this assignment when I diagnosed a case study with the disorder of Tem.
4. I was confident in this assignment when I diagnosed a case study with the disorder of Pliq.
5. I was confident in this assignment when I diagnosed a case study with the disorder of Hix.
6. I was confident in this assignment when I diagnosed a case study with the disorder of Wos.

Real Label Condition

1. I was confident in this assignment when I diagnosed a case study with the disorder of Obsessive Compulsive Disorder.
2. I was confident in this assignment when I diagnosed a case study with the disorder of Anxiety.
3. I was confident in this assignment when I diagnosed a case study with the disorder of Schizophrenia.
4. I was confident in this assignment when I diagnosed a case study with the disorder of Bipolar.
5. I was confident in this assignment when I diagnosed a case study with the disorder of Dissociative Identity Disorder.
6. I was confident in this assignment when I diagnosed a case study with the disorder of Depression.

Appendix H

Association Check for Novel Label Condition

1) Which disorder below does Duv correspond with:
    a. Don't know
    b. Anxiety
    c. Obsessive Compulsive Disorder
    d. Depression
    e. Dissociative Identity Disorder
    f. Schizophrenia
    g. Bipolar
    h. Other

2) Which disorder below does Baj correspond with:
    a. Don't know
    b. Anxiety
    c. Obsessive Compulsive Disorder
    d. Depression
    e. Dissociative Identity Disorder
    f. Schizophrenia
    g. Bipolar

3) Which disorder below does Tem correspond with:
    a. Don't know
    b. Anxiety
    c. Obsessive Compulsive Disorder
    d. Depression
    e. Dissociative Identity Disorder
    f. Schizophrenia
    g. Bipolar

4) Which disorder below does Pliq correspond with:
    a. Don't know
    b. Anxiety
    c. Obsessive Compulsive Disorder
    d. Depression
    e. Dissociative Identity Disorder
    f. Schizophrenia
    g. Bipolar

5) Which disorder below does Hix correspond with:
    a. Don't know
    b. Anxiety
    c. Obsessive Compulsive Disorder
    d. Depression
    e. Dissociative Identity Disorder
    f. Schizophrenia
    g. Bipolar

6) Which disorder below does Wos correspond with:
    a. Don't know
    b. Anxiety
    c. Obsessive Compulsive Disorder
    d. Depression
    e. Dissociative Identity Disorder
    f. Schizophrenia
    g. Bipolar

Appendix I

Distraction Task

Directions: Answer the following math equations by typing your answer in the box below.

547 – 3 =

544 – 3 =

541 – 3 =

538 – 3 =

535 – 3 =

532 – 3 =

529 – 3 =

526 – 3 =

523 – 3 =

520 – 3 =

517 – 3 =

514 – 3 =

511 – 3 =

508 – 3 =

505 – 3 =

Appendix J

Debriefing

Real Label Condition.

The study you just participated in is examining the effects of the presentation of material on inductive learning and retention. You time and willingness to participate in this study is greatly appreciated. If you feel you may be experiencing adverse reactions due to this study please speak to with the researcher or contact Ms. Sadie Lisenby, Director of Counseling at 901.321.3527 or slisenby@cbu.edu.

If you have any questions about this study or would like information on the results please contact Chanda Murphy at 901.321.3338 or cmurphy6@cbu.edu. Also please email Chanda Murphy if you would like access to the study program used in this experiment to further study the differences between the disorders.

Thank you for your participation.

Novel Label Condition

The study you just participated in is examining the effects of the presentation of material on inductive learning and retention. You time and willingness to participate in this study is greatly appreciated. If you feel you may be experiencing adverse reactions due to this study please speak to with the researcher or contact Ms. Sadie Lisenby, Director of Counseling at 901.321.3527 or slisenby@cbu.edu.

If you have any questions about this study or would like information on the results please contact Chanda Murphy at 901.321.3338 or cmurphy6@cbu.edu. Also please email Chanda Murphy if you would like access to the study program used in this experiment to further study the differences between the disorders.

Thank you for your participation.