The Pennsylvania State University

The Graduate School

College of Education

IMPACT OF INSTRUCTIONAL MATERIALS ELICITING LOW AND HIGH COGNITIVE LOAD ON SELF-EFFICACY AND DEMONSTRATED KNOWLEDGE

A Dissertation in

Educational Psychology

by

Whitney Alicia Zimmerman

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2015

The dissertation of Whitney Alicia Zimmerman was reviewed and approved * by the following:

Jonna M. Kulikowich
Professor of Education
Dissertation Co-Adviser
Co-Chair of Committee

P. Karen Murphy
Professor of Education
Dissertation Co-Adviser
Co-Chair of Committee

Stephanie L. Knight
Professor of Education

Gary Kuhne
Associate Professor of Education

Robert Stevens
Co-Program Coordinator, Educational Psychology Program

*Signatures are on file in the Graduate School.

Abstract

The impact of instructional materials designed to elicit two varying levels of perceived cognitive load on post-secondary introductory statistics students' self-efficacy, perceived knowledge, and demonstrated topic knowledge was examined. The chi-square test of independence was the focal topic. Lower and higher levels of perceived cognitive load were elicited using worked example and conventional problem solving methods, respectively. Participants were randomly assigned to the worked example or conventional problem solving condition. First, all participants completed pre-test measures of self-efficacy, perceived knowledge, and demonstrated knowledge. Self-efficacy and demonstrated knowledge were measured using researcher-developed scales while perceived knowledge was measured using two items in which participants were asked to rate their perceived knowledge of introductory statistics and the chi-square test of independence. Then participants watched an instructional video and completed three practice problems. Those who were assigned to the worked example group studied one worked example and then completed two conventional problems. Those who were assigned to the conventional problem solving group completed three conventional problems. Finally, participants completed post-test measures of self-efficacy, perceived knowledge, and demonstrated knowledge along with a demographic questionnaire. There were increases in self-efficacy, perceived knowledge of the chi-square test of independence, and demonstrated knowledge from the pre- to post-instruction administrations however there were no significant differences between the worked example and conventional problem solving groups. There was not a significant change in perceived knowledge of introductory statistics nor was there a difference in the group means. The correlations between demonstrated knowledge and each of the other variables were also examined for the worked example and conventional problem solving groups. Residual gain scores and Fisher's Z transformation were used to compare changes in the correlations between self-efficacy and demonstrated knowledge in the two groups. The worked example group saw a greater correlation

between residual gains than the conventional problem solving group (r=.717, p<.001, n=51; r=.489, p<.001, n=49, respectively). The difference between the correlations in the worked example and conventional problem solving groups was not statistically significant (z=1.777, p=.076). Given the results, more research on the relations between self-efficacy and experiences is needed before conclusions can be made.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

Students' behaviors are influenced by their perceptions of a learning task. Even the brightest students may struggle when faced with a relatively simple task if they perceive themselves as incapable in relation to the demands of the task. Both the learners' perceptions of their own abilities and their perceptions of the task relate to how they view the task and their behaviors.

In educational psychology research, individuals' perceptions of their abilities to successfully perform given tasks, also known as self-efficacy, have been studied by social cognitivists such as Bandura (1977b, 1997). Individuals' perceptions of the cognitive resources utilized when completing a task are referred to as perceived cognitive load which has been studied by educational psychologists such as Sweller (2010a, 2010b) and Paas (Paas, Renkl, & Sweller, 2003; Paas, van Gog, & Sweller, 2010). The relations between self-efficacy and cognitive load, however, have not yet been studied in great detail.

The present research study examined the relations between demonstrated topic knowledge, perceived topic knowledge, cognitive load, and self-efficacy in the domain of statistics at the post-secondary level. The chi-square test of independence was chosen as the specific topic of interest. Statistics was selected because it is a domain with which many students struggle and sometimes attempt to avoid (Gordon, 2004; Onwuegbuzie & Wilson, 2003). Any findings that could influence how statistics is taught could have a great impact on many students' experiences and learning. The chi-square test of independence was chosen as the topic of focus because it is an introductory test that is not typically taught until later in the course from which participants were recruited. The chi-square test of independence requires only a minimal level of prerequisite knowledge of statistics and involves relatively basic mathematical operations. Additionally, its basic procedures can be taught in a relatively short time. These characteristics all made the chi-square test of independence an ideal topic to be used in the present research studies.

The following sections introduce the primary constructs of interest. Each section defines the construct and briefly summarizes related research. In chapter two, the existing literature is reviewed in greater detail. First, knowledge is reviewed with an emphasis on demonstrated topic knowledge and perceived topic knowledge. Second, cognitive load is reviewed followed by self-efficacy. Finally, the topic of interest, the chi-square test of independence is introduced.

Knowledge

The definition of knowledge that will be used throughout this dissertation is "an individual's personal stock of information, skills, experiences, beliefs, and memories" (Alexander, Schallert, & Hare, 1991, p. 317). In chapter two, types of knowledge will be discussed, namely declarative, procedural, and conditional knowledge. The present study examined demonstrated topic knowledge of the chi-square test of independence, which is a combination of declarative, procedural, and conditional knowledge of the topic. Demonstrated topic knowledge is influenced by experiences, such as being exposed to a related reading (e.g., Garner & Gillingham, 1991; Murphy & Alexander, 2004).

Perceived topic knowledge was also examined in the present dissertation research. This construct refers to an individual's judgment about the information that he or she knows about a given subject area (Murphy & Alexander, 2004). According to Murphy and Alexander, measures of perceived knowledge require individuals to think about their knowledge which is a type of metacognition. Perceived knowledge is different from demonstrated topic knowledge in that it is an individual's perception as opposed to an observable behavior. Like demonstrated topic knowledge, perceived topic knowledge can be influenced by experiences; individuals with more pertinent experiences tend to have higher levels of perceived topic knowledge (e.g., Dodson & Lewallen, 2011; Hsiao, Van Riper, Lee, Chen, & Lin, 2011). In addition, perceived topic knowledge tends to increase following an experience with the topic, such as reading an article (e.g., Murphy & Alexander, 2004) or completing a course (e.g., Rutledge, Siebert, Chonody, & Killian, 2011).

Cognitive Load

In addition to examining self-efficacy, the present study examined the implications of instructional materials designed to elicit varying levels of perceived cognitive load. The cognitive demands associated with any given mental activity can be described as an interplay of intrinsic, extraneous, and germane cognitive loads (Sweller, 2010a). According to Sweller, intrinsic and extraneous cognitive load can be described in terms of element interactivity where an element is defined as a unit of knowledge. Element interactivity is defined as a measure of total number of elements that must be processed simultaneously, that is, the units cannot be separated and processed independently (Paas et al., 2010). More difficult tasks involve higher levels of element interactivity (Sweller, 2006, 2010b). Intrinsic cognitive load refers to the element interactivity required by the complexity of the content that is to be learned (Sweller, 2010b). Extraneous cognitive load is that which is imposed by factors other than the content to be learned, such as the instructional method (Plass, Moreno, & Brünken, 2010b; Sweller, 2010b). A third type of cognitive load is also often described as germane cognitive load which is different from intrinsic and extraneous cognitive loads in that it does not impede on working memory, instead it has a positive impact on learning (Paas & van Gog, 2006; Plass et al., 2010b). Sweller (2010b) described germane cognitive load as “purely a function of the working memory resources devoted to the interacting elements that determine intrinsic cognitive load” (p. 126). The most effective instruction minimizes extraneous cognitive load and increases germane cognitive load (Sweller et al., 1998). Together, intrinsic, extraneous, and germane cognitive loads can be used to describe a learning scenario.

Since the 1980s, the study of cognitive load has grown to become a major theoretical construct in the field of educational psychology (Paas et al., 2003). Cognitive load has been studied in a wide variety of educational contexts, including mathematics (e.g., Paas & van Merriënboer, 1994; Sweller & Cooper, 1985), statistics (e.g., Leppink, Broers, Imbos, van der Vleuten, & Berger, 2012; Paas, 1992), and educational technology (e.g., Kalyuga, 2012; Sweller, 2008). Cognitive load theory is based on the premise that working memory has a limited capacity (Sweller, 2010b). If the cognitive resources required to complete a task are too great, in other words, more than working memory can simultaneously process, individuals will experience cognitive overload (Ayres & Paas, 2012).

Cognitive load has great practical implications in the field of education, particularly in the design of instructional materials which are often examined in terms of extraneous cognitive load (Low, Jin, & Sweller, 2011; Plass et al., 2010b). According to Sweller (2010b), "nonoptimal instructional procedures are referred to as imposing an extraneous cognitive load. Cognitive load theory is primarily concerned with techniques designed to reduce extraneous cognitive load" (p. 125).

Instructors can influence students' perceived cognitive load via their instruction. Instructional materials can be designed to decrease or increase learners' perceived cognitive load (e.g., Ayres, 2006a; Paas, 1992). Worked examples are often used to reduce cognitive load during problem solving activities in well-structured domains (e.g., Moreno, Reisslein, & Ozogul, 2009; Paas, 1992; Renkl & Atkinson, 2010). There have been many studies comparing the use of worked examples to conventional problem solving including one study performed by Sweller and Cooper (1985) which compared the use of conventional problem solving and worked examples in secondary and post-secondary algebra students. In Sweller and Cooper's study, learners in the worked example condition solved problems in less time and with fewer errors than those in the conventional problem solving condition.

Paas (1992) studied the impact of different instructional methods by comparing the use of conventional problem solving, worked examples, and completion problems in secondary-level statistics students. The content of the materials in Paas' study focused on measures of central tendency. Following general instruction, participants in each treatment group were given instructional materials.

The problems that each group was presented with were the same, however, the instructional method varied. The conventional problem solving group received all unsolved problems that they needed to solve themselves. The worked examples group received problems that were already solved followed by a similar problem they had to solve. The completion problems group was given partly-worked problems to complete followed by a similar problem that they had to solve on their own. Participants who were presented with the worked examples required the least amount of time to study. Those who received the conventional problems scored the lower than the other groups on the measures of near and far knowledge transfer. Perceived mental effort when completing problems was higher for the conventional conditional when compared to the worked example and completion groups.

Self-Efficacy

According to Bandura (1977b, 1997), self-efficacy is an individual's perceptions of his or her abilities to execute a given task to a certain level of success. He states, "perceived self-efficacy is concerned not with the number of skills you have, but with what you believe you can do with what you have under a variety of circumstances" (1997, p. 37). The task-specific nature of self-efficacy distinguishes it from other, broader constructs such as self-concept or self-esteem (Bandura, 1997; Pajares, 1996). Self-efficacy is often cited as being influenced by four sources: personal experiences, watching others, psycho-physical status, and feedback from others (Schunk, 2012). In the present study, the focus will be on personal experiences with instructional materials as an influence on self-efficacy.

Self-efficacy has been studied in a variety of contexts from physical activity (Bandura & Cervone, 1983; Jung & Brawley, 2011) to substance abuse (e.g., Connor, George, Gullo, Kelly, & Young, 2011; Senbanjo, Wolff, Marshall, & Strang, 2009). In statistics education research, self-efficacy has been studied both in terms of self-efficacy for learning statistics and self-efficacy for performing statistics (Cashin, 1999, 2001; Finney & Schraw, 2003). Relations between self-efficacy and domain knowledge have been studied in the domain of statistics (e.g., Finney & Schraw, 2003; Lane, Hall, & Lane, 2004; Zare, Rastegar, & Hosseini, 2011). Finney and Schraw found that course grades were positively correlated with statistics self-efficacy as measured at the end of the course (r = .496); self-efficacy measured at the beginning of the course was also positively correlated with final course grades, though the correlation was not as strong (r = .340). Changes in statistics self-efficacy over time have also been investigated (e.g., Finney & Schraw, 2003; Lane et al., 2004). For example, undergraduate students' statistics efficacy increased on average by approximately two standard deviations from the beginning to the end of one introductory statistics course [t (109) = 18.64, p <.001; Finney & Schraw, 2003].

The findings of self-efficacy research have been consistent across contexts in that self-efficacy and behaviors are related (Bandura, 1997). Individuals with higher levels of self-efficacy were more likely to engage in related behaviors (Pajares, 1996). Individuals with higher levels of self-efficacy were also more likely to persist when faced with a challenge (Lent, Brown, & Larkin, 1984; Schunk, 1981). The relations between self-efficacy and observable abilities has also been studied and is known as self-efficacy calibration (Alexander, 2013; Glenberg, Sanocki, Epstein, & Morris, 1987). Self-efficacy is of great interest to educational researchers because if its connections with motivation, persistence, performance, and self-regulation (Bandura, 1982; B. J. Zimmerman, 2000).

Chi-Square Test of Independence

In the present study, within the domain of statistics, the chi-square test of independence was selected as the topic of focus. The chi-square test of independence statistically tests the null hypothesis that two or more categorical variables are unrelated to one another; that is, they are independent. The alternative hypothesis is that the categorical variables are related to one another in some way. In the present study, only 2-way chi-square tests of independence were used for simplicity of computations.

As an example of a 2-way chi-square test of independence, one could test the independence of gender (man, woman) and whether or not one consumes coffee (yes, no). This would be a 2x2 chi-square test of independence because there are two levels of the gender variable and two levels of the coffee consumption variable. If the null hypothesis is rejected, then there is statistical evidence that gender and coffee consumption are related. Knowing an individual's gender can help you predict whether or not that individual consumes coffee, and vice versa (Coladarci, Cobb, Minium, & Clarke, 2011; Dodge, 2008). If the null hypothesis is not rejected, then there is not sufficient statistical evidence to state that there is a relationship between the two variables.

The chi-square test of independence is described in greater detail in Chapter 2. In the following sections, gaps in the literature and the purpose of the present research is discussed. The two pilot studies and the full study are outlined. Finally, a list of relevant definitions is given.

Gap in Literature

The instructional materials and the design of the present dissertation research were based on the work Sweller and Cooper (1985) and Paas (1992). Unlike Paas' research, the present study only compared worked examples to conventional problem solving without a completion condition. Also, time was not a variable of interest in the present study. The variables of interest were self-efficacy, perceived knowledge, and demonstrated knowledge which were not all examined in the aforementioned studies. As in Paas' study, the domain of interest in this dissertation was introductory statistics. The present study examined the impact of worked example versus conventional problem solving instructional materials on self-efficacy, perceived knowledge, and demonstrated knowledge as pertaining to the topic of the chi-square test of independence.

A limited amount of research has studied cognitive load and self-efficacy in conjunction with one another (e.g., Vasile, Marhan, Singer, & Stoicescu, 2011; Zheng, McAlack, Wilmes, Kohler-Evans, & Williamson, 2009). For example, Zheng et al.'s experimental study of problem solving found that learners who were given materials that were perceived as less cognitively demanding (i.e., lower cognitive load) had higher levels of self-efficacy when compared to participants who were presented with materials that elicited higher cognitive load. In an observational study, Vasile and colleagues found that academic self-efficacy and working memory capacity were moderately, positively related; while working memory capacity is not a direct measure of perceived cognitive load, working memory capacity may impact an individual's perceptions of the cognitive demands of a task. There is a gap in the literature in the examination of self-efficacy and cognitive load. More specifically, there is need for research examining the relations between self-efficacy for learning a specific topic and cognitive load while completing a task on that topic.

Purpose

The purpose of this dissertation research was to examine self-efficacy, perceived knowledge, and demonstrated knowledge in relation to varying levels of perceived cognitive load induced by different instructional materials. There were two primary research questions:

1. To what extent do instructional materials designed to elicit low and high levels of cognitive load impact self-efficacy, perceived knowledge, and demonstrated knowledge?
2. To what extent do instructional materials designed to elicit low and high levels of cognitive load impact the relations between self-efficacy, perceived knowledge, and demonstrated knowledge?

Outline of Studies

This dissertation research consisted of a series of three studies; there were two pilot studies which informed one larger, full study. The first pilot study served to design and validate measures of self-efficacy, perceived knowledge, and demonstrated knowledge. These measures were all designed concerning the topic of the chi-square test of independence. The psychometric properties of each of these instruments were examined and changes were made as deemed necessary in accordance with the results of the first pilot study.

The second pilot study served to design and validate the instructional materials that would be used to elicit low and high levels of cognitive load. All participants were presented with the same instructional video which introduced the topic of the chi-square test of independence. Participants were then randomly assigned to receive practice problems designed to elicit low or high levels of cognitive load. These materials were designed following a review of the work of Sweller and Cooper (1985) and Paas (1992). A worked example was used to elicit lower levels of cognitive load. Conventional problem solving was used to elicit higher levels of cognitive load. Participants were asked to rate their perceived cognitive load when using these instructional materials.

The full study used the instruments and instructional materials developed in the pilot studies. Participants completed the measures of self-efficacy, perceived knowledge, and demonstrated knowledge before and after being presented with the instructional materials. Again, participants were randomly assigned to receive either the instructional materials designed to elicit low levels of cognitive load or the instructional materials designed to elicit high levels of cognitive load. The ability of the instructional materials in eliciting different levels of perceived cognitive load was evaluated. Changes in self-efficacy, perceived knowledge, and demonstrated knowledge were evaluated in relation to the instructional materials. And, changes in the relations between these variables at the pre- and post-instruction administrations were compared for the worked example and conventional problem solving conditions.

Limitations of Sample

A limitation of this research was the samples that were employed. The samples utilized in the two pilot studies were relatively small and homogeneous in terms of a number of variables. The sample of the first pilot study consisted entirely of graduate students enrolled in education-related programs. The sample of the second pilot study consisted entirely of undergraduate students enrolled in two introductory-level statistics courses.

In the full study, all participants were enrolled in the same instructor's introductory statistics course. The response rate within that course was relatively low. There were 312 students enrolled in the course and 103 students participated in the research study for a response rate of 33.0%. Students were offered extra credit in their introductory statistics course in exchange for their participation in the study. It is unknown whether or not the sample was representative of the class in which recruitment occurred.

The sample in the full study was homogeneous in terms of other variables as well. All participants were traditionally aged students between the ages of 18 and 24. In terms of age only, there were no adult learners in the sample. These limitations effect the generalizability of the results.

Definitions of Key Terms

The following terms will be used often in the present research. These terms will be discussed in greater detail in Chapter 2. Here, brief definitions of key terms are given.

Knowledge, in the present studies, will be defined as "all that a person knows or believes to be true, whether or not it is verified as true in some sort of objective or external way" (Alexander, Schallert, & Hare, 1991, p. 317).

Topic knowledge is information known about a given focal subject. In the present research, the topic of interest is that of the chi-square test of independence.

Demonstrated topic knowledge is information known about a given focal subject that can be observed by others.

Perceived topic knowledge refers to an individual's judgment of the amount of information he or she knows about a given subject (Murphy & Alexander, 2004).

Cognitive load theory describes the cognitive demands of a mental activity as being a combination of intrinsic, extraneous, and germane cognitive loads (Sweller, 2010b).

Intrinsic cognitive load refers to the mental demands of a learning task that can be attributed to the complexity of the information to be learned (Sweller, 2010b).

Extraneous cognitive load refers to the mental demands of a learning task that are imposed by factors other than the content to be learned, such as the methods through which the content is presented to the learner (Plass et al., 2010b; Sweller, 2010b).

Germane cognitive load has been described by Sweller (2010b) as "purely a function of the working memory resources devoted to the interacting elements that determine intrinsic cognitive load" (p. 126).

Perceived cognitive load is an individuals' personal judgment of the amount of mental effort that he or she is expending during the completion of a given task (Paas, 1992). It is often measured using Paas' self-report scale, a 9-point scale ranging from a score of one signifying "*very, very low mental effort*" to a score of nine signifying "*very, very high mental effort*."

Self-efficacy is an individual's perceptions of his or her abilities to execute a given task to a certain level of success (Bandura, 1977a, 1977b, 1978, 1982, 1997, 2007).

Chapter 2: Review of Literature

The following review of literature examined previous research on the constructs of knowledge, self-efficacy, and cognitive load. In each section, the construct at hand was defined and a general overview was given. When available, research featuring the domain statistics was included. In each section, the measurement of each construct was discussed. Each section is concluded with an overview of themes within the research with a focus on relations with the other variables of interest when possible. After reviewing each of the literature pertaining to knowledge, self-efficacy, and cognitive load, the relations between self-efficacy and cognitive load were described with an emphasis on gaps in the existing literature.

This review of literature was constructed following a series of searches of Penn State University Libraries' search engine: LionSearch (Penn State University Libraries, 2014). Through LionSearch, the majority of the Libraries' resources have been compiled and can be searched simultaneously. This search engine allowed the researcher to limit results to pieces from peer-reviewed publications. In some cases, books with historical value or those that provided experts' overviews of research in the field were also reviewed (e.g., Bandura, 1977b, 1986; Plass, Moreno, & Brünken, 2010a).

Knowledge

The construct of knowledge has been studied for centuries from a variety of perspectives. Within the field of educational psychology, it is often defined in terms of what it is purporting to measure (Murphy & Mason, 2006). In this review, one generally accepted definition of knowledge will be applied: "knowledge refers to an individual's personal stock of information, skills, experiences, beliefs, and memories" (Alexander, Schallert, & Hare, 1991, p. 317). Knowledge is often discussed in terms of form (i.e., disciplinary, domain, or topic; Alexander, et al., 1991) and type (i.e., declarative, procedural, and conditional; Leonard, 2002).

Demonstrated topic knowledge. In the current study, demonstrated topic knowledge is the form of knowledge of interest. More specifically, demonstrated topic knowledge concerning the topic of the chi-square tests of independence. This topic serves as the content for the instructional materials and is the focus of the demonstrated knowledge measures as well as the self-efficacy and perceived knowledge measures. Whenever possible, examples concerning the topic of chi-square tests of independence will be given.

Types. The declarative, procedural, and conditional classifications of knowledge will be used throughout this discussion of topic knowledge. Alexander et al. (1991) reminds us, "it is important to remember that these three types of knowledge are distinct; the acquisition of knowledge in one form does not automatically and immediately guarantee knowledge in the other forms" (p. 323). The following sections will describe the differences between these types of knowledge.

Declarative knowledge. Declarative knowledge generally refers to knowledge of explicit understandings or facts. It is information that can be declared or stated verbally (Alexander et al., 1991). Within the topic area of the chi-square test of independence, an example of declarative knowledge could be knowledge of the formula for computing the test statistic $\left(\chi^2 = \sum \frac{(O-E)^2}{E}\right)$. The ability to state this equation from memory would be classified as declarative knowledge. Note, this is only the stating of the equation. The derivation of the equation or the application of the equation would require additional pieces of knowledge that would likely include a combination of declarative, procedural, and conditional knowledge types.

Another example of declarative knowledge related to the topic of the chi-square test of independence is knowledge of the assumptions of a statistical test. If students are asked to state three assumptions of a chi-square test of independence, then their ability to recall that these assumptions include (a) a random sample, (b) independent observations, and (c) nominal-level data is an example of declarative knowledge. These are facts or propositions that individuals can state.

Procedural knowledge. Procedural knowledge is knowledge concerning how to carry out given actions (Leonard, 2002). It is “knowledge of certain processes or routines” (Alexander et al., 1991, p. 323). While beginners often need to refer back to pieces of declarative knowledge when performing a procedure, experts can often complete procedures quickly with no need to reference other pieces of knowledge. Procedural knowledge can be gained without learners being aware of its acquisition (Stadler, 1989).

There are many examples of procedural knowledge in action in the domain of statistics. Using a table, such as a chi-square table, to look up critical values is one example of procedural knowledge. If students are asked to find the critical value given a chi-square test of independence with four degrees of freedom and a significance level of .05, they will need use their procedural knowledge. On the chi-square table they will need to locate the column associated with the .05 significance level. Then, they will need to locate the row associated with four degrees of freedom. Finally, they find where the appropriate column and row intersect to identify 9.488 as the critical value.

Procedural knowledge will be the focus of the instructional materials in the proposed study. The intervention will provide instruction on the procedures associated with performing a chi-square test of independence. The majority of items on the demonstrated knowledge measure will require participants to draw from their procedural knowledge.

Conditional knowledge. The third type of knowledge is conditional knowledge. This is knowledge of when to use certain types of declarative or procedural knowledge (Alexander et al., 1991; Bruning, Schraw, Norby, & Ronning, 2004). In other words, the conditions under which one should apply other types of knowledge.

Conditional knowledge is applied often in real-life situations involving statistics. Research questions and the types of data available create the conditions under which certain statistical tests are appropriate and some are inappropriate. If given a cross-tabulation of frequency counts and asked if the categorical variables are related to one other, students would need to use conditional knowledge to select the correct analysis. They would need to evaluate the conditions of the problem to determine that a chi-square test of independence is the appropriate test to use in this situation.

Summary. Students studying the topic of the chi-square test of independence use declarative, procedural, and conditional knowledge. In solving a statistical problem in real life, all three types of knowledge are applied at various moments. For example, if given a dataset and asked to determine if two independent variables are related, students would first need to examine the data and use their declarative and procedural knowledge to determine that they are working with independent frequency counts. They could then use that knowledge to determine that under those conditions a chi-square test of independence would be an appropriate analysis to perform which would be an application of their conditional knowledge. The calculation of the test statistic requires the use of procedural knowledge. A beginner may also need to recall declarative knowledge before being able to perform a procedure. For example, they may first need to think about and write down the declarative knowledge of the formula before performing their calculations. Finally, once the test statistic has been calculated, conditional knowledge of when to reject and when to fail to reject the null hypothesis must be applied. This process, as a whole, of conducting a chi-square test of independence requires the use of declarative, procedural, and conditional knowledge.

Experiences and demonstrated topic knowledge. Demonstrated topic knowledge is influenced by experiences. Demonstrable topic knowledge is typically greater after the presentation of related information. This increase is often studied by comparing test scores before and after a related intervention (e.g., Garner & Gillingham, 1991; Murphy & Alexander, 2004). Increases in demonstrated topic knowledge over time are also evidence of the connection between demonstrated topic knowledge and experiences (Pinter, Matchock, Charles, & Balch, 2014).

Garner and Gillingham (1991) measured individuals' demonstrated knowledge on the topic of Stephen Hawking before and after presenting them with a related reading. Participants were undergraduate students recruited from an intermediate psychology course. They were randomly assigned to one of two groups. Both groups received a text to read. One groups' reading included a paragraph that was interesting but unnecessary while the second group did not receive that paragraph. There was not a significant difference for the two groups in terms of demonstrated topic knowledge assessed using unstructured [$t(160) = -0.70$, $p = .24$] or structured [$t(160) = -0.13$, $p = .45$] recall measures. Though different measures of topic knowledge were used pre- and post-intervention, the researchers did note that pre-intervention topic knowledge was low and after post-intervention knowledge was high as measured by unstructured and structured recall. This suggests that demonstrated topic knowledge may have increased following the related reading.

Similarly, in a study of persuasive texts, Murphy and Alexander (2004) measured demonstrated topic knowledge before and after participants read articles on the topics of physician-assisted suicide, AIDS vaccinations, and racial integration in schools. The vast majority of participants were junior and senior undergraduate students enrolled in educational psychology, child psychology, and child development courses. The correlation between pre- and post-test demonstrated topic knowledge was positive and strong ($r = .662$). A statistically significant increase in demonstrated topic knowledge was observed from pre- to post-test with a large effect size [$F(1, 233) = 94.70$, $p < .0001$, $\eta^2 = .30$]. Changes were different for the three articles [$F(2, 466) = 10.0$, $p < .001$, $\eta^2 = .09$]. Thus, providing evidence for the impact of experience on demonstrated topic knowledge and also suggesting that its impact may vary on the basis of other factors, such as the topic of the article in this case.

The impact of experiences on demonstrated knowledge is also evident in the research of Pinter et al. (2014). They compared the demonstrated knowledge of three groups of undergraduate students: senior psychology majors, freshman intending to declare psychology as their major, and freshmen nonpsychology majors who scored in the top ten percent of their introductory psychology class. All freshmen were students enrolled in an introductory psychology course. Participants completed ETS's Psychology Major Field Test and a short-answer essay assessment. There were statistically significant group differences in terms of both Psychology Major Field Test results [$F(2, 90) = 15.92$, $p < .001$, $\eta^2 = .26$] and the short-answer essay scores [$F(2, 78) = 13.62$, $p < .001$, $\eta^2 = .26$]. For both variables, the graduating senior psychology majors outperformed both freshmen groups. Though their measures of demonstrated knowledge were broad, the results suggest that more experiences lead to greater demons tread knowledge.

The research presented here provides evidence that demonstrated knowledge increases after individuals have related experiences. That related experience can be a relatively short reading, such as in Murphy and Alexander's (2004) or Garner and Gillingham's (1991) research. In Pinter et al. (2014) groups with varying levels of experience (i.e., freshmen who had completed one psychology course and graduating senior psychology majors) were compared. The group with the most experience had the highest test scores.

While Pinter et al. (2014) used an existing measure (ETS's Psychology Major Field Test), the other studies reviewed here used more task-specific measures. Both Garner and Gillingham (1991) and Murphy and Alexander (2004) used instruments that were directly in line with the texts that they used as their interventions. The following section describes the measurement of demonstrated knowledge used in these studies in greater detail.

Measurement of demonstrated topic knowledge. Demonstrated knowledge can be measured using multiple choice or open-ended questions. When open-ended items are used, multiple raters are needed to assure reliability of scores (e.g., Garner & Gillingham, 1991; Murphy & Alexander, 2004). Here, the items and scoring schemes used in Garner and Gillingham (1991) as well as in Murphy and Alexander (2004) will be discussed.

To measure knowledge on the topic of Stephen Hawking, Garner and Gillingham (1991) presented participants with a series of open-ended questions. They used five questions, such as “Who is Stephen Hawking?” and “What current theories exist about the origins of the universe?” (p. 313). Participants’ responses were rated by two individuals. A three point scale was used: 0 = inaccurate or blank; 1 = incomplete; and 2 = accurate and complete. In their first round of rating, a .98 agreement rate was achieved. Following the initial round of rating, the scorers discussed and resolved the items on which they disagreed.

Murphy and Alexander (2004) also used open-ended questions to measure participants’ topic knowledge relating to assigned articles. For each article, participants were given a list of key points and asked to write what they knew about each point. A six point scale was used: 0 = no response or a completely incorrect response; 1 = limited amount of at least partially correct information; 2 = limited amount of information but was correct; 3 = extended responses that were mostly correct; 4 = extended responses that were completely correct; and 5 = correct extended responses that included information from the assigned article. Overall, there were three raters who all rated the same random sample of ten percent of papers for which an inter-rater agreement level of .97 was computed.

The methods of Garner and Gillingham (1991) and Murphy and Alexander (2004) are similar in that both used open-ended items. The former employed a three point scale while the later employed a six point scale. Both had very high levels of rater agreement.

Demonstrated topic knowledge can be studied through the use of multiple choice or open-ended questions. Both of the studies examined here used open-ended questions. The high levels of rater agreement are indicative of relative objectiveness. That is, an individual with some expertise on the topic can rate the work of others.

In the following section, perceived topic knowledge will be discussed. Perceived topic knowledge is an individual's own judgment of how much he or she knows about a given topic. Thus, it is a perception of topic knowledge.

Perceived topic knowledge. According to Murphy and Alexander (2004), "because the perceived knowledge measure requires respondents to self-judge their background knowledge, it is more reflective of participants' metacognitive knowledge" (p. 344). Thus, perceived knowledge is different from demonstrated topic knowledge in that it is an individual's perception as opposed to an observable behavior. It is also different from self-efficacy in that it is topic-specific as opposed to task-specific. The following sections summarize research concerning perceived knowledge in terms of its influences and its measurement.

Experiences and perceived topic knowledge. Experiences can influence perceptions of topic knowledge in a number of ways. Perceived knowledge tends to increase over time with learning experiences (Dodson & Lewallen, 2011; Hsiao et al., 2011; Murphy & Alexander, 2004; Rutledge et al., 2011). Here, research that examined the impact of experiences on perceived knowledge is reviewed.

In a study of undergraduate nursing students' perceived knowledge of genetics, senior students reported the highest levels of perceived knowledge followed by juniors, sophomores, and, with the lowest reported levels of perceived knowledge, freshmen [F (3, 259) = 5.574, $p < .001$; Dodson & Lewallen, 2011]. While Dodson and Lewallen's research used a sample of nursing students in the United States, Hsiao et al. (2011) performed a similar study with a sample of nursing students in Taiwan. A significant effect for educational level was also observed with the mean scores for freshmen and sophomores being lower than they were for juniors and seniors ($\overline{\mathrm{X}} = 136.31$, 137.92, 153.31, and 152.2 for freshmen, sophomores, juniors, and seniors, respectively). As students progress from their freshman to the senior years of college, their perceived knowledge increase on average.

Murphy and Alexander (2004) compared perceived knowledge of three different topics (physician-assisted suicide, AIDS vaccinations, and racial integration in schools) before and after participants read articles on the topics. Participants were primarily junior and senior undergraduate students enrolled courses in educational psychology, child psychology, and child development. Participants' perceived knowledge increased on average from pre- to post-test with a large effect size [F (1, 233) = 161.58, $p < .0001$, $\eta^2 = .41$]. Changes in perceived knowledge were different for the three articles [F (2, 466) = 14.05. $p < .001$, $\eta^2 = .12$], however perceived knowledge did increase for each topic. Again, this provides evidence for the impact of experiences on individuals' perceived knowledge.

The impact of experiences was also examined in a study by Rutledge, Siebert, Chonody, and Killian (2011) of undergraduate and graduate students' perceived knowledge of human sexuality before and after courses in the topic area. At the beginning of the course students were asked to rate their knowledge of human sexuality on a five point scale from 1 = *not knowledgeable at all* to 5 = *completely knowledgeable*; the average rating was 3.2 with a standard deviation of 0.90. In this study, at the end of the course students were asked to rate how much they knew before the class (i.e., "retrospective pretest"; p. 482). Here, the average rating was 2.9 with a standard deviation of 0.86. Students perceived their earlier ratings as being overestimates after they completed the course as evidenced by the significant change in perceived pre-course knowledge [t (533) = 3.82, $p < .001$, $d = 0.339$]. At the end of the course students were also asked to rate their current perceived knowledge. The average rating was 3.9 with a standard deviation of 0.64. Current perceived knowledge increased from the beginning to the end of the course [t (533) = 9.73, $p < .001$, $d = 0.864$].

Experiences can impact perceived knowledge in a number of ways. As undergraduate students progress from freshman year to senior year, their perceived knowledge tends to increase (e.g., Dodson & Lewallen, 2011; Hsiao et al., 2011). Perceived knowledge also increases as a result of being exposed to new information such as through a course (e.g., Rutledge et al., 2011) or a reading (e.g., Murphy &

Alexander, 2004). Rutledge et al. (2011) also noted that learners' perceived knowledge may be overestimated prior to instruction. Or, their perceptions of knowledge may change as a result of experiences, such as that which is gained through formal educational experiences.

The previously reviewed studies examined the impact of related experiences on students' perceived knowledge. Research in social psychology has examined the impact of related versus unrelated experiences (e.g., Tormala & Petty, 2007). This is of particular interest because it may explain some previously unexplained variance in perceived knowledge self-report ratings.

In a study of persuasion, Tormala and Petty (2007) presented participants with information about two stores: "Smith's" and "Brown's." Participants were first presented with either a low or high level of information about Smith's; they were then all presented with the same amount of information about Brown's. The individuals who were first presented with a little information about Smith's had a higher level of perceived knowledge about Brown's than those who were first presented with a high level of information about Smith's [$t\ (42) = 3.46$, $p = .001$]. The participants who were first presented with a little information about Smith's also had more positive attitudes toward Brown's [$t\ (42) = 3.32$, $p = .002$]. This shows that recent learning experiences can impact individuals' perceived knowledge ratings.

In a follow-up study, Tormala and Petty (2007) examine the impact of being presented with unrelated information. They compared the impact of being presented information about a store, as they did in the previous study, to the impact of being presented with information about an unrelated topic: automobiles. There was not a significant main effect for the type of information provided (i.e., store versus automobile) on perceived knowledge [$F\ (1, 76) = 0.25$, $p = .62$]. There was still a main effect for amount of information [$F\ (1, 76) = 10.26$; $p < .002$]. Participants who were presented with little information on the first topic had higher levels of perceived knowledge on the second as opposed to those who received a high amount of information on the first topic. Thus showing that even information in unrelated areas can impact individuals' perceptions of their knowledge.

Measurement of perceived knowledge. Perceived knowledge is measured using self-report methods. Most often, participants are asked to rate how much they know about a given set of topics (e.g., Maradiegue et al., 2005; Tormala & Petty, 2007). Thus, the level of measurement is typically at the topic level, in contrast to self-efficacy which is measured at the more specific task level. Self-efficacy will be discussed in greater detail in a later section.

A number of studies from the field of nurse education were reviewed. They all used similar methods for measuring perceived knowledge. Dodson and Lewallen (2011) and Hsiao et al. (2011), for example, studied perceived knowledge in nursing students by presenting participants with surveys derived from one initially developed by Maradiegue et al. (2005). The instrument consisted of 39 questions concerning knowledge related to genetics terms, such as "mitosis" and "Trisomy 21," to which participants rated their level of knowledge as "no," "minimal," "some," or "high" for each individual topic.

None of the reviewed studies from nurse education (i.e., Dodson & Lewallen, 2011; Hsiao et al., 2011; Lahl et al., 2013; Maradiegue et al., 2005) included a measure of reliability. However, Lahl et al. (2013) did discuss the validity of their scale. After its initial development, the scale was reviewed by 10 nurses with related expertise to provide evidence for content validity.

While self-report measures of perceived topic knowledge seem to be overwhelmingly the most used method, the scales that have been used vary somewhat. In nurse education, Maradiegue et al.'s (2005) scale has been used a number of times. The scale uses a four point scale that ranges from no knowledge to high knowledge with a list of 39 related terms. Lahl et al. (2013) also used a four point scale, although with different anchors: no confidence through great confidence. Note this later measure of perceived knowledge actually focuses on confidence level. Perceived knowledge and confidence levels seem to be used interchangeably, particularly in the nurse education articles cited here.

Tormala and Petty (2007), in their persuasion research, used a nine point scale. Their anchors differed depending on the item. The item that is most applicable to the proposed research study is that of "How knowledgeable do you feel you are about Brown's Department Store?" which was anchored with "*not knowledgeable at all"* at the lower bound and "*extremely knowledgeable*" at the upper bound. Scores from this three item scale had a very high reliability coefficient (Cronbach's alpha = .94).

A five point scale was used in the study of human sexuality by Rutledge et al. (2011). Participants rated a single item that ranged from 1 = *not at all knowledgeable* to 5 = *very knowledgeable*. Because a single item was used, no measure of internal consistency was available. Similarly, Murphy and Alexander (2004) used a 10 point scale that asked participants how much they knew about three topics ranging from *relatively nothing* to *a great deal*. Results from the three items resulted in a Cronbach's alpha of .65, however it is important to note that the three topics were diverse, concerning doctor-assisted suicide, the creation of an AIDS vaccine, and racial integration in schools, thus the internal consistency of the responses to these items may not be expected to be more than moderately high.

Summary. Knowledge may be described using the schema that distinguishes between declarative, procedural, and conditional knowledge (see Alexander et al., 1991; Leonard, 2002). Demonstrated topic knowledge and perceived topic knowledge each contribute unique information when discussing learning. Demonstrated topic knowledge is learners' observable declarative, procedural, and conditional knowledge. While learners may possess other information, it is the observable knowledge that is of interest. This is often measured using written assessments (e.g., Garner & Gillingham, 1991; Murphy & Alexander, 2004).

Perceived topic knowledge is a broad measure of learners' perceptions of the information that they possess concerning a topic. This is in contrast to self-efficacy which is a task specific perception that will be discussed in greater detail in a later section. In a review of literature concerning perceived knowledge, the impact of experiences emerged as a theme. Experiences often lead to increases in perceived knowledge (Dodson & Lewallen, 2011; Hsiao et al., 2011). However, experiences may also impact how individuals perceive their level of knowledge (e.g., Rutledge et al., 2011; Tormala & Petty, 2007).

There is much research that links perceptions of cognitive effort (i.e., cognitive load) to the acquisition of knowledge. The following section will examine research on cognitive load in greater detail. A number of frequently cited phenomena related to cognitive load will be explored in greater detail.

Cognitive Load

Cognitive load research describes the cognitive demands of a mental activity as being a combination of intrinsic, extraneous, and germane cognitive loads. According to Sweller (2010a), "all instructional material imposes a working memory or cognitive load, and that cognitive load can be divided into two independent categories – intrinsic and extraneous – with a third category, germane cognitive load, dependent on intrinsic cognitive load" (p. 40). Intrinsic and extraneous loads can be described in terms of element interactivity. An element is a unit of knowledge that is required for learning. Element interactivity is a measure of total number of elements that must be processed simultaneously. That is, they cannot be separated and processed separately. More difficult tasks require greater amounts of element interactivity. The level of element interactivity required by a task is not only dependent on the task, but also on characteristics of the learner. Individuals with greater levels of expertise are able to combine elements using their existing schemas (Sweller, 2006, 2010a, 2010b).

In terms of element interactivity, intrinsic cognitive load refers to the element interactivity required by the content to be learned. That it, is the cognitive load that is caused by the complexity of the material (Sweller, 2010a, 2010b). Extraneous cognitive load is imposed by factors other than the content to be learned, such as the instructional method (Plass et al., 2010b; Sweller, 2010b). For example, in an online statistics module on the topic of measures of central tendency, intrinsic cognitive load is that which is necessary to learn the content that is being presented: measures of central tendency. Extraneous cognitive load is that which is not directly related to the content of central tendency. In this case, extraneous cognitive load may include the cognitive demands of using a computer or navigating the course management system.

A third type of cognitive load is also often described: germane cognitive load. Germane is different from intrinsic and extraneous loads in that it is not considered to be a strain on working memory (Plass et al., 2010b; Sweller et al., 1998; Sweller, 2010b). Rather, germane cognitive load has a positive impact on learning as it “reflects the effort that contributes to the construction of schemas” (Sweller et al., 1998, p. 259). Sweller (2010b) describes germane cognitive load as “purely a function of the working memory resources devoted to the interacting elements that determine intrinsic cognitive load” (p. 126). Recent research has begun to explore ways to increase germane cognitive load. Paas and van Gog (2006), for example, suggest that prior knowledge should be taken into account when discussing methods for increasing germane load. Learners with differing levels of prior knowledge may benefit from different strategies.

Cognitive load research is typically focused on instructional methods associated with optimal use of cognitive resources (Mayer, 2013). Much of the research has been conducted in controlled laboratory settings and over shorter periods of time in which the impact of motivation may be consistent throughout the data collection period (Brünken, Plass, & Moreno, 2010; Paas, Tuovinen, van Merriënboer, & Darabi, 2005). While these laboratory-based research studies give us evidence to support cognitive load theory, they examine learning in a very controlled setting as opposed to a real-life learning setting. Paas et al. (2005) suggest that research should move to non-laboratory environments that allow for more authentic evaluation of the application of the theory.

Variables related to cognitive load. In this section three often cited factors that influence cognitive load and can be used to manipulate perceived cognitive load will be reviewed: task difficulty, prior knowledge, and instructional methods. Through this review of previous research, the positive and negative aspects of using method of manipulating cognitive load in research will be discussed.

Task difficulty. Cognitive load can be manipulated by changing the difficulty of the content being taught or the problems being solved (Beckmann, 2010; Leahy & Sweller, 2008). The difficulty of the content can also be manipulated through instructional methods. For example, introducing learners to isolated elements as opposed to whole tasks (Ayres, 2006a). A number of research studies that successfully and unsuccessfully manipulated cognitive load via task difficulty are reviewed here.

Beckmann (2010) proposed a framework for testing the intrinsic and extraneous cognitive loads associated with tasks. Intrinsic cognitive load was increased by making tasks more difficult. This was done by manipulating the number of steps required to complete the tasks, more steps were associated with higher intrinsic cognitive load. Participants in the condition that required fewer steps outscored the participants in the condition that required more steps [F (2, 456) = 135.84, $p < .01$, partial $\eta^2 = .37$].

Ayres (2006a) studied the impact of reducing cognitive load by altering the complexity of the content being taught. In the first of a series of studies, a sample of eighth grade females studying algebra was randomly assigned to one of two groups. Both groups were presented with a set of worked problems. The isolated group received worked examples that were broken down into subsections in which each required one calculation. The integrated group received complete worked examples not separated into smaller elements. Students who were presented with the worked examples featuring isolated elements made fewer errors [t (32) = 2.16, $p < .05$, Cohen's $d = 0.8$] and reported lower levels of cognitive load [t (32) = 2.11, $p < .05$, Cohen's $d = 0.7$]. Thus providing evidence that cognitive load can be impacted by the difficulty of the tasks presented.

In a second study, again using a sample of eighth grade females, Ayres (2006a) compared three groups: one group was presented with a single element at a time; a second group was presented with whole task, in other words, all elements together; and, a third group was presented with a combination of tasks that progressed from isolated elements to whole tasks. Participants were also classified as being above or below average in the sample in terms of their mathematical abilities. There was not a significant treatment by ability interaction [F (2, 70) = 1.05, $p = .36$]. Again, participants in the group that received the content as single elements gave the lowest ratings on the perceived cognitive load measure [F (2, 70) = 5.59, $p <.01$, $d = 0.8$]. A main effect for ability was observed. Individuals who were classified as above average mathematics ability gave lower ratings of perceived cognitive load [F (2, 70) = 7.64, $p < .01$, $d = 0.6$].

Leahy and Sweller (2008) examined the impact of low versus high element interactivity along with the impact of imagination in a problem solving activity for third grade students. Element interactivity served as a within subjects independent variable and imagination served as a between subjects independent variable. Participants were asked to respond to 17 items: 12 high element interactivity items and 5 low element interactivity items. While they did observe a significant imagination by element interactivity interaction [F (1, 28) = 7.04, $p = .01$] and a main effect of the imagination condition [F (1, 28) = 5.51, $p = .002$], there was not a statistically significant effect of element interactivity [F (1, 28) = 1.19, $p = .28$]. This led the researchers to conclude that the items were not of different difficulty levels. In an examination of the cell means, for the learners in the imagination group, scores were higher on the high element activity items. For learners in the non-imagination group, scores were higher on the low element activity items. No measures related to perceived cognitive load were included.

Manipulating task difficulty has been shown to have an impact on cognitive load (e.g., Ayres, 2006a; Beckmann, 2010). However, when examining the relationship between cognitive load and demonstrated knowledge, the content that is presented may impact the knowledge that can be gained. Thus, altering task difficulty is not an appropriate method for manipulating cognitive load when demonstrated knowledge is also an outcome variable of interest. Also, task difficulty may be affected by the knowledge or experiences of the participants as shown by Leahy and Sweller (2008).

Similar to Ayres' (2006a) approach of altering the way information is presented, some researchers have manipulated task difficulty by providing background information to learners (e.g., Rey & Fischer, 2013). Doing so does not change the task at hand, rather it changes the learners' perceptions of the difficulty of the task. The same task may be difficult to novices but easy to individuals with more related experiences. In the following section, the impact of prior topic knowledge on cognitive load is discussed.

Prior topic knowledge. Prior related knowledge is one factor that influences the cognitive load necessary to complete a task. According to Kalyuga, Ayres, Chandler, and Sweller, (2003), "learning reduces working memory limitations by enabling the use of schemas, stored in long-term memory, to process information more efficiently" (p. 23). That is, the presence of prior topic-relevant knowledge reduces the processing needs required for the performance of a given task. By reducing some of these processing needs, working memory can be allocated for additional tasks.

In a study of learning from online hypertext, Amadieu, van Gog, Paas, Tricot, and Mariné (2009) examined the impact of prior knowledge on adults' abilities to learn from online concept maps with hierarchical and network structures. The researchers state, "learners' prior knowledge is important because it allows coping with the cognitive demands imposed by hypertext learning since learners can use their own mental representation of the knowledge domain to guide their navigation and the processing of the content of the concept map" (p. 377). Cognitive load was higher for individuals with lower levels of prior knowledge when answering questions concerning factual and conceptual knowledge during the posttest phase [Mann-Whitney U (12, 12) = 40, p = .027; Mann-Whitney U (12, 12) = 39, p = .025 for factual and conceptual knowledge items respectively]. However, there were not significant differences between the low and high prior knowledge groups in terms of mental effort while completing the learning task. Through the use of eye tracking methods, however, the researchers were able to find that individuals with high prior knowledge were able to focus their attention on material relevant for learning when compared to the low prior knowledge group. Thus showing that there are differences in the ways that individuals with varying levels of prior knowledge interact with content.

The impact that prior topic knowledge has on cognitive load is also demonstrated by the expertise reversal effect. Prior topic knowledge can have an impact on the effectiveness of instructional techniques. That is, methods that work well with learners who have very little related knowledge may not be effective when learners have greater levels of related knowledge. The expertise reversal effect has been examined in a number of previous studies (for reviews see Kalyuga et al., 2003; Kalyuga & Renkl, 2010) its practical applications have also been explored (see Kalyuga, Rikers, & Paas, 2012; Kalyuga, 2007).

In the previously discussed work of Ayres (2006a) in which eighth grade algebra students received problems either by single element, by whole problem, or as a progression from single element to whole problem, the expertise reversal effect could be seen in the analysis of errors made during testing. Above average participants in the single element group, while giving the lowest cognitive load ratings, made the most errors. Above average participants in the whole problem group gave the highest ratings of cognitive load but made the least errors. The opposite occurred in the below average group in which the fewest errors were made by participants in the single element group and the most errors were made by those in the whole problem and progression groups. For those participants in the single element group, which resulted in the lowest ratings of perceived cognitive load, the below average students outperformed the above average students. These results suggest that the relationship between perceived cognitive load and demonstrated ability may be impacted by individuals' prior knowledge.

Leppink, Broers, Imbos, van der Vleuten, and Berger (2012) examined the expertise reversal effect in the domain of statistics. Their study used introductory-level content related to the sampling distribution of the mean. Learning was compared for participants with low and high levels of prior knowledge when exposed to four different types of instruction: (1) reading, (2) open-ended questions, (3) open-ended questions that require arguments, and (4) worked examples of arguments (i.e., they were given the answers to the open-ended questions with arguments). The results showed that for learners with low levels of prior knowledge, performance was best when open-ended questions (without arguments) or worked examples were used while for learners with high levels of prior knowledge, performance was best when open-ended questions with or without arguments were used. Under all conditions, participants with high prior knowledge outperformed those with low prior knowledge. These results show that prior knowledge impacts post-intervention performance and that the impact of different instructional methods varies depending on learners' levels of prior knowledge. Thus, it is important for researchers to take into account participants' topic knowledge prior to intervention.

Recently, Rey and Fischer (2013) examined the expertise reversal effect by randomly assigning participants to "novice" and "expert" conditions. Their study also used content related to statistics. All participants were given instructional materials. The expert group was given additional expository examples and illustrations. That is, expertise was defined level of knowledge obtained during the learning exercise. Participants were also randomly assigned to two groups: one with and one without explanations. A statistically significant interaction was found between level of knowledge (i.e., novices versus expert) and explanations (i.e., presence or absence of) on the dependent measure of transfer [$F(1, 89) = 5.68, p < .05, \eta_p^2 = .06$]. For the novice group, the explanations had a positive effect [$p = .06; d = 0.46$] while for the expert group, the explanations had a negative effect [$p < .05, d = 0.55$]. However, these statistically significant results were only present for the dependent measure of transfer and not for the dependent measure of retention [$F(1, 89) = 0.53, p = .47$]. These results provide evidence that it is possible to replicate the expertise reversal effect using experimental methods, thus providing further validity for the phenomena. The results are particularly convincing as the study was of a relatively short duration. The mean time spent on the intervention was between 13 and 16 minutes.

All of the aforementioned reviews and experiments support the notion that prior knowledge has an impact on cognitive load. Prior knowledge allows working memory to process information more efficiently (Kalyuga et al., 2003; Sweller, 1976). It also impacts the effectiveness of instructional methods which is known as the expertise reversal effect (Ayres, 2006a; Kalyuga & Renkl, 2010; Kalyuga, 2007; Leppink et al., 2012; Rey & Fischer, 2013).

Instructional methods. Instructional methods are often linked to extraneous cognitive load. Inefficient instructional methods contribute to extraneous cognitive load (Plass et al., 2010b; Sweller, 2010b). According to Sweller (2010b), "Cognitive load theory is primarily concerned with techniques designed to reduce extraneous cognitive load" (p. 125). Research on cognitive load has been used to guide best practice in teaching in a wide variety of disciplines including statistics (e.g., Lovett & Greenhouse, 2000).

There are a number of well-studied, instructionally-related effects that can be applied to reduce or increase cognitive load. Some examples of instructionally-related effects include the worked example effect, goal-free effect, completion effect, redundancy effect, split-attention effect, and modality effect (Beckmann, 2010). Here, the worked example and redundancy effects will be examined in greater detail.

Worked example effect. Worked examples are typically described in three parts: (1) the stated problem, (2) the solution steps, and (3) the final solution (Moreno et al., 2009; Renkl, 1997). Their purpose is to decrease extraneous cognitive load by giving learners the steps as opposed to making the learners solve the problems themselves. Worked examples are often used in well-structured problem-solving domains such as physics, computer programming, and statistics with novice learners (Moreno, Reisslein, & Ozogul, 2009; Renkl & Atkinson, 2010).

There have been many studies comparing the use of worked examples to conventional problem solving. One of the earliest studies is that of Sweller and Cooper (1985). They compared the use of conventional problem solving and worked examples in high school and college algebra students. They hypothesized that participants in the worked examples condition would outperform those in the conventional problem solving condition because “worked examples may potentially be more apt to result in schema acquisition than the solution of conventional problems” (p. 69). Procedurally, they first gave all participants worked examples. Any questions that the participants had at that point were answered. Then, participants were given a sheet of problem sets. For the worked examples group, the first problem of each set was worked through. Then, all participants were given an assessment of their demonstrated knowledge as related to the problem sets they had just solved. During this assessment period, participants did not have access to any of their previously solved problems.

In Sweller and Cooper's (1985) study, participants in the worked examples group needed less time to study [Mann-Whitney U (10, 10) = 25] and were able to complete similar problem in less time than students in the conventional problem solving group [Mann-Whitney U (11, 11) = 27.5]. They also produced fewer errors [Mann-Whitney U (11, 11) = 30]. This study did not include a measure of participants’ perceived cognitive load.

In a different study on cognitive load in statistics education, Paas (1992) measured cognitive load in conjunction with the use of worked examples. He compared the use of conventional problem solving, worked examples, and completion problems in a statistics course for 16-18 year old learners. The content of the materials focused on measures of central tendency. To begin, all participants were given general instruction on the topic of measures of central tendency as well as an explanation of the measure of cognitive load that they would be using.

Following the general instruction, participants in each group were given specific instructional materials. The problems that each group was presented with were the same, however the instructional method was different. The conventional problem solving group received all unsolved problems that they needed to solve themselves. The worked examples group received problems that were already solved followed by a similar problem they had to solve. The completion problems group was given partly-worked problems to complete followed by a similar problem that they had to solve on their own (Paas, 1992).

In Paas' (1992) study, participants who were given worked examples required the least amount of time to study [F (2, 39) = 6.86, $p < .01$; post-hoc statistics not given]. The worked examples and completion groups outperformed the conventional problem solvers [F (2, 39) = 13.55, $p < .001$; post-hoc statistics not given]. Scores for the worked examples condition were higher on measures of both near- and far-transfer [post-hoc comparison statistics were not given]. Perceived mental effort when completing problems was highest for the conventional conditional when compared to the worked example and completion groups [F (2, 39) = 17.79, $p < .01$; post-hoc statistics not given]. While this study did include a measure of cognitive load, the measure was used primarily in conjunction with problem solving activities as opposed to instructional activities.

There is ample research showing that worked examples, when appropriately written, can reduce cognitive load and in turn improve learning. Worked examples reduce extraneous cognitive load by not requiring learners to perform procedures themselves. Instead, learners can focus on aspects of the examples. Worked examples have been used in well-defined domains such as statistics (e.g., Paas, 1992). Kalyuga (2012) points out that the effectiveness of worked examples may be detrimental to learners with advanced knowledge as they “may cause these learners to corefer and reconcile these solution steps with their available procedural knowledge structures, thus imposing additional extraneous cognitive load” (p. 186). He does, however, go on to suggest the use of worked examples with learners with inadequate prior knowledge.

Redundancy effect. The inclusion of redundant information in instructional materials increases extraneous cognitive load (Sweller, Ayres, & Kalyuga, 2011b). This phenomenon is known as the redundancy effect. This effect has been shown to be related to demonstrated topic knowledge (e.g., Pastore, 2012) and perceptions of cognitive load (e.g., Pociask & Morrison, 2004).

Pastore (2012) examined the redundancy effect in conjunction with time-compressed instructional materials. Redundancy was induced by presenting information in auditory and written forms. Identification and comprehension were measured. There was not a significant redundancy by time-compression interaction effect. There was a significant main effect for redundancy in terms of measures of both identification [$F (1, 148) = 18.57$, $p < .001$, partial $\eta^2 = .111$] and comprehension [$F (1, 148) = 6.24$, $p = .014$, partial $\eta^2 = .040$]. For both dependent variables, the group that did not receive redundant information scored better than the group that did receive redundant information.

Pastore (2012) also measured perceived cognitive load. A seven-point self-report measure was used. After the instruction and before the post-test, participants used the scale to rate how easy or difficult they perceived the instructional materials to be. Using this method, there was not a significant difference between the ratings given by participants in the redundant and non-redundant groups [$F (1, 121) = 1.93$, $p = .167$, partial $\eta^2 = .016$].

Pociask and Morrison (2004) examined the redundancy effect and the split-attention effect in relation to demonstrated topic knowledge and demonstrated psychomotor performance in the domain of physical therapy was examined. Two groups were compared, a non-integrated group in which redundant information was presented, and an integrated group in which redundant information was eliminated. The content the both groups were presented with was equivalent. Participants in the group that received redundant instructional materials scored lower on measures of demonstrated topic knowledge [F (1, 39) = 16.564, $p < .001$, $\eta^2 = .30$] and psychomotor performance [F (1, 39) = 29.15, $p < .001$, $\eta^2 = .43$]. Participants in the non-integrated group (i.e., redundant group) also had significantly higher levels of perceived cognitive load immediately after the instruction [F (1, 39) = 6.02, $p = .019$, $\eta^2 = .13$] and immediately following the assessment of psychomotor skills [F (1, 39) = 7.76, $p = .008$, $\eta^2 = .17$]. This provides evidence that perceptions of cognitive load can be manipulated through the use of different instructional materials.

Pociask and Morrison's (2004) study showed a moderate effect of redundancy on perceived cognitive load. They found statistically and practically significant differences between the participants who received redundant and non-redundant instructional materials both after instruction and after the assessment of their demonstrable skills. However, Pastore's (2012) research failed to find a practically significant impact of redundancy on perceived cognitive load.

Summary of manipulating cognitive load. Cognitive load can be manipulated in a number of ways. From the previously reviewed pieces, research has shown that cognitive load is related to task difficulty (e.g., Ayres, 2006a; Beckmann, 2010), individuals’ prior knowledge (e.g., Amadieu et al., 2009; Kalyuga et al., 2003; Leppink et al., 2012; Rey & Fischer, 2013) and instructional methods (e.g., Paas, 1992; Pociask & Morrison, 2004; Sweller & Cooper, 1985). In terms of manipulating cognitive load experimentally, previous researchers have been able to do so by employing methods related to all three of these aspects.

Manipulating cognitive load via differences in task difficulty impacts the intrinsic cognitive load of the activity. However, it is important to first provide evidence for the differences in difficulty, as seen by the findings of Leahy and Sweller (2008) who did not observe significant differences between items that they identified as being low and high in element interactivity. In research on teaching and learning, altering content impacts the validity of comparing perceived cognitive load with demonstrated topic knowledge. The demonstrated topic knowledge measure should be in line with the information presented. This cannot be done if participants are presented with different information.

Manipulating participants' prior knowledge may be done on a micro level as demonstrated by Rey and Fischer (2013). Prior knowledge on a macro level and other related individual characteristics such as related coursework, aptitude, and all other previous experiences cannot practically be manipulated in an experimental study. Some have used pre-existing differences in knowledge as a variable in research (e.g., Amadieu et al., 2009; Ayres, 2006a; Leppink et al., 2012).

In educational research examining demonstrated knowledge in conjunction with cognitive load, the manipulation of extraneous cognitive load via the use of different instructional techniques is a common approach. In the previous section, two well-researched instructional manipulations of extraneous cognitive load were reviewed: the worked examples effect and the redundancy effect. Worked examples have been shown to be related to increases in learning efficiency and decreases in perceived cognitive load (e.g., Paas, 1992). However, the effectiveness of worked examples may be influenced by individuals' levels of prior knowledge (e.g., Kalyuga, Chandler, & Sweller, 2000), as seen by the expertise reversal effect (see Kalyuga et al., 2003; Kalyuga & Renkl, 2010).

Research concerning the impact of redundancy on perceived cognitive load has mixed results. Pociask and Morrison (2004) observed a moderate effect of redundancy on participants' perceived cognitive load both immediately following instruction and immediately following an assessment of their demonstrable skills. Pastore (2012) did not observe a statistically or practically significant effect of redundancy on learners' perceived cognitive load.

Given the findings of previous studies, the favorable method for manipulating cognitive load depends on the other variables of interest and characteristics of the population being studied. For populations with little prior knowledge, worked examples can be used to decrease cognitive load effectively. For populations with prior knowledge, however, the effectiveness of worked examples may be impacted by the expertise reversal effect.

Measurement of cognitive load. While a variety of methods have been used to measure cognitive load such as reaction time (e.g., Schoor, Bannert, & Brünken, 2012), eye tracking (e.g., Wang, Yang, Liu, Cao, & Ma, 2014; Zheng & Cook, 2012), and physiologically-based methods (e.g., Uysal, 2013), self-report measures, specifically the nine-point scale used by Paas (1992), are preferred by many researchers: "it has a great advantage in that it requires very little time to administer and does not interfere with the instructional interventions that are being investigated" (Ayres & Paas, 2012, p. 827). Self-report measures have also been touted as being "more sensitive and far less intrusive than the physiological measure" (Sweller, Ayres, & Kalyuga, 2011, p. 73). Paas' (1992) self-report measure of cognitive load is one of the most frequently cited methods. A criticism of this method is that it does not distinguish between the different sources of cognitive load because it provides a singular measure of overall cognitive load (Ayres, 2006b).

In terms of the specific aspects of self-report measures, van Gog, Kirschmer, Kester, and Paas (2012) examined the impact of the timing and frequency of measuring mental effort. From a series of four studies, they found that results differed depending on when and how many times participants were asked to rate their mental efforts. They conclude that it is preferable to administer mental effort measures after each individual task as opposed to waiting until after all tasks have been completed.

In the study that examined cognitive load while solving introductory statistics problems, Paas (1992) introduced a self-report method for measuring cognitive load. The 9-point scale ranges from a score of one signifying "*very, very low mental effort*" to a score of nine signifying "*very, very high mental effort*." This scale was used to measure participants' perceived mental effort after each task. In this original study, reliability was measured using Cronbach's alpha ($\alpha = .90$).

Zheng and Cook (2012) used a combination of eye tracking, reaction time, accuracy analyses, and the Cognitive Load Questionnaire (from Paas, 1992) which asked participants three questions: (1) "In solving the preceding problems, I invested: (1 - *very low cognitive effort* to 9 - *high cognitive effort*)"; (2) "I experienced the foregoing problem solving as: (1 – *not difficult at all* to 9 – *very difficult*)"; and (3) "How easy/difficult was the problem solving to understand? (1 – *very easy* to 9 – *very difficult*)." The independent variable was the presence or absence of graphics in instructional materials. Theoretically, the inclusion of graphics should reduce cognitive load needs. The response time, accuracy analyses, and Cognitive Load Questionnaire self-report scores showed no statistically significant differences between the graphic present and absent conditions. The results of the analyses of eye tracking data did show differences in terms of pupil diameter [$F\ (1, 46) = 3.74$, $p = .059$, $\eta^2 = .075$] and peak amplitude [$F\ (1, 46) = 3.35$, $p = .074$, $\eta^2 = .068$], however effect sizes were relatively small. Eye tracking data was also used to analyze accumulated cognitive load by examining the area under the response curve and the results were statistically significant with a medium effect size [$F\ (1, 46) = 10.63$, $p = .002$, $\eta^2 = .188$]. Data collected using the eye tracking equipment and the Cognitive Load Questionnaire were not significantly correlated with one another. The researchers suggest that "these different measures of cognitive load may actually be measuring separate aspects of cognitive load" (p. 243).

Ayres (2006b) used a rating scale similar to that of Paas (1992). In a study that held extraneous cognitive load constant and focused on changes in intrinsic cognitive load, a 7-point self-report scale was used. The scale ranged from “extremely easy” to “extremely difficult.” The measure was found to be highly reliable. In a study in which participants used this scale after answering each of 16 items, a Cronbach’s alpha of .90 was computed for the scores. In a second study in which participants used the scale after answering each of 32 items, a Cronbach’s alpha of .97 was computed for scores.

The results of the studies utilizing self-report measures by Paas (1992) and Ayres (2006b) provide evidence for the reliability of scores obtained through this method. However, the aspect of cognitive load that these instruments are measuring has been questioned by Zheng and Cook (2012) as well as by Ayres (2006b). Because the proposed study is primarily concerning with learners’ perceptions, self-report methods of measuring cognitive load have been selected. Future studies may further examine how other methods capture different aspects of cognitive load.

Self-Efficacy

Self-efficacy is an individual’s perceptions of his or her abilities to execute a given task to a certain level of success. It is task specific. Self-efficacy is not what an individual can do, rather it is what an individual perceives him or herself to be capable of (Bandura, 1977a, 1977b, 1978, 1982, 1997, 2007).

Research in a variety of disciplines has provided evidence for a positive correlation between self-efficacy and behaviors (Bandura, 1982; Schunk, 2012; B. J. Zimmerman, 2000). Learners with higher levels of self-efficacy are more likely to choose to engage in behaviors that will lead to better performance, such as studying or putting forth extra effort when working on a class project. Individuals with higher levels of efficacy are more likely to persist when faced with a challenge. The relations between self-efficacy and observable abilities is referred to as self-efficacy calibration (Alexander, 2013; Glenberg et al., 1987). Self-efficacy is of great interest in the field of education because if its connects with motivation, persistence, performance, and self-regulation (Bandura, 1982; Schunk & Usher, 2011; Schunk, 2012; B. J. Zimmerman, Bandura, & Martinez-Pons, 1992; B. J. Zimmerman, 2000).

Variables related to self-efficacy. In general, self-efficacy is influenced by four primary sources: modeling/vicarious experiences, personal experiences, psycho-physical status, and feedback from others (Schunk, 2012). Modeling and vicarious experiences occur when the learner observes someone else attempt a task. Efficacy increases as the leaner perceives that it is possible to succeed (e.g., Bandura & Jourden, 1991; Bandura, Reese, & Adams, 1982; Bartsch, Case, & Meerman, 2012; Ozer & Bandura, 1990).

Self-efficacy is influenced by personal experiences with success or failure. That is, individuals attempt a task, they succeed, and therefore they believe that they can be successful again in the future (e.g., Locke, Frederick, Lee, & Bobko, 1984). The impact of experiences on self-efficacy and the correlation between self-efficacy and demonstrable knowledge will be discussed in the following section. Specifically, research in teacher education that has identified changes in the relationship between self-efficacy and demonstrated following experiences will be discussed as this is an area in which these relations have been explored in the past (i.e., Woolfolk Hoy & Spero, 2005; W. A. Zimmerman, Parker, & Knight, 2011).

Psycho-physiological states can impact efficacy. For example, depression and anxiety can have a negative impact on efficacy. Physical states, such as illness or tiredness, can also impact an individual's perception of what he or she is capable of at the time (Davidson, 2003; Schunk, 2012).

Finally, feedback from others can also alter an individual's perceptions (e.g., Collins, Bandura, 1992). Encouragement from others can increase efficacy. Feedback from others can impact an individual's efficacy, and thus behaviors, regardless of the accuracy of the feedback (e.g., Weinberg, Gould, & Jackson, 1979).

Moderating effect of experience. Research in the area of teacher education has shown that experiences can have a moderating effect on the relations between self-efficacy and demonstrable knowledge. For example, Woolfolk Hoy and Spero (2005) measured teaching efficacy in individuals at the beginning of their teacher education program, at the end of their student teaching experience, and after one year of working full-time as a teacher. They observed an increase in teaching efficacy from the beginning of the study to the end of student teaching. However, they observed a decrease in efficacy from the end of student teaching to the end of the first year of full-time teaching. The researchers attribute this change to the lesser amount of support that teachers had available during their first year of independent teaching as opposed to their time as a student in the teacher education program. Thus, self-efficacy decreased as the perceived demands of teaching increased.

Similar observations were made by Zimmerman, Parker, and Knight (2011) who examined changes in teaching efficacy as a result of a professional development experience. They found that for teachers who participated in the professional development experience there was a negative relationship between teaching efficacy and observable teaching behaviors. For teachers who did not participate in the professional development experiences there was a positive relationship between teaching efficacy and observable teaching behaviors. The learning experience changed how teachers perceived their abilities.

Both the research of Woolfolk Hoy and Spero (2005) and Zimmerman, Parker, and Knight (2011) are examples of the interrelationships between self-efficacy, demonstrable knowledge, and experience. They both provide evidence for the impact of experiences on self-efficacy. The latter research also ties in demonstrable knowledge in the form of observable teaching behaviors. There is still much to be learned about the moderating effect of experiences on the relationship between self-efficacy and demonstrable knowledge.

The present dissertation studies will examine learning using an introductory statistics topic. Thus, the following section will review self-efficacy specifically in the area of statistics. This will be followed by a review of the literature concerning self-efficacy for learning in online environments as the current study will employ online instructional methods.

Statistics self-efficacy. Because self-efficacy is task specific, the following will review research specifically on self-efficacy for statistics-related tasks. This will include research on self-efficacy for performing statistical tasks as well as self-efficacy for learning statistics. First, related research studies will be summarized. Then, the measurement of statistics self-efficacy will be discussed.

***Research*.** Research concerning self-efficacy for statistics-related tasks has examined its relationship with demonstrated knowledge (e.g., Finney & Schraw, 2003; Lane, Hall, & Lane, 2004; Zare, Rastegar, & Hosseini, 2011). Changes in statistics self-efficacy over time have also been examined (e.g., Finney & Schraw, 2003; Lane et al., 2004). The findings suggest that a positive relationship exists between statistics self-efficacy and demonstrated knowledge and that self-efficacy tends to increase with related experiences.

Finney and Schraw (2003) studied both self-efficacy for learning statistics and current self-efficacy for performing specific statistics-related tasks at the beginning and end of an undergraduate-level introductory statistics course. They developed two parallel instruments for measuring students' self-efficacy for learning statistics and current statistics self-efficacy. These instruments are examined in greater detail in the discussion of measuring statistics self-efficacy which follows. Tasks on their scale included "Identify the factors that influence power" and "Distinguish between a population parameter and a sample statistic." Demonstrated knowledge was measured using final course grades and a post-test measure consisting of items related to each of the items on the self-efficacy scales. Course grades and post-test scores were both most highly correlated with current statistics self-efficacy measured at the end of the course (r = .496, r = .441, respectively). Self-efficacy for learning statistics at the beginning of the course was also significantly positively correlated with course grades and post-test scores (r = .340, r = .229, respectively). The correlations between current self-efficacy and knowledge that were both measured at the end of the course were stronger than the correlations between self-efficacy for learning at the beginning of the course and measures of demonstrated knowledge at the end of the course.

Zare, Rastegar, and Hosseini (2011) examined the ability of statistics self-efficacy, anxiety, and achievement goals to predict academic achievement in Iranian statistics university students. The instrument developed by Finney and Schraw (2003) to measure self-efficacy for learning statistics was used. A relatively strong positive association was observed between self-efficacy for learning statistics and final grades ($r = .47$). The correlation between final grades and self-efficacy for learning statistics was stronger than that of the correlation between grades and anxiety ($r = -.41$) or grades and a measure performance avoidance goals ($r = -.31$). Of the variables examined in this study, self-efficacy for learning statistics was the strongest predictor of demonstrated knowledge as measured by final course grades.

In a study of undergraduate sports study majors, Lane, Hall, and Lane (2004) also examined the relationship between self-efficacy and performance in a statistics course. Like Finney and Schraw (2003), Lane et al. measured statistics self-efficacy at the beginning of the course and later in the course (the seventh week). Their operationalization of statistics self-efficacy was broader than that of Finney and Schraw's. They examined six aspects related to success in a statistics course: statistical theory, lecture behavior, information technology, motivated behavior, managing time, and general competencies. The association between self-efficacy at the beginning of the course and at the seventh week was strong ($r^2 = .43$). The relationship between self-efficacy at the beginning of the course and performance on a later module was not statistically significant. However, the relationship between self-efficacy later in the course and performance on the module at around the same time was statistically significant and moderately strong ($r^2 = .17$). Again, this shows that the correlation between self-efficacy and demonstrated knowledge is strongest when both are measured with close temporal proximity.

Changes in statistics self-efficacy over time have been examined in a number of studies. In Finney and Schraw's (2003) research, students' statistics efficacy increased on average by approximately two standard deviations from the beginning to the end of their statistics course [t (109) = 18.64, p <.001]. While Finney and Schraw's study found significant increases in statistics self-efficacy over time, Lane et al. (2004) did not find evidence for an increase in statistics self-efficacy over a time period of approximately seven weeks [t (57) = 0.42, p = .34], however their operationalization of statistics self-efficacy included a broader range of tasks.

Measurement of statistics self-efficacy. Within the domain of statistics, self-efficacy has been operationalized in a number of ways. For example, some researchers have focused on students' perceptions of their current abilities to answer statistics problems (Finney & Schraw, 2003). Others have examined self-efficacy for successfully completing a statistics course (Bartsch et al., 2012; Lane et al., 2004), self-efficacy for using a specific instructional method (Hall & Vance, 2010), or self-efficacy for learning statistics (Chiesi, Primi, & Carmona, 2011; Finney & Schraw, 2003; Zare et al., 2011).

Regardless of the perspective that one is coming from, all of these operational definitions rely on self-report methods of data collection. Typically, participants are asked to rate their confidence in their abilities to complete give tasks. The tasks are dependent on how the researcher is defining statistics self-efficacy. The rating scales used vary. Bartsch et al. (2012) used a ten point scale (1 = *no confidence*, 10 = *total confidence*). Finney and Schraw (2003) used a six point scale (1 = *no confidence at all*, 6 = *complete confidence*). Lane et al. (2004) used a five point scale (0 = *not at all confident,* 4 = *very confident*).

The aforementioned study by Finney and Schraw (2003) included measures of both self-efficacy to learning statistics (SELS) and current statistics self-efficacy (CSSE). The SELS and CSSE items were identical though the instructions were slightly different in that participants were asked either "rate your confidence in ***learning***" (p. 184) or "rate your confidence in your ***current ability***" (p. 182), all other aspects of the instruments were the same. Despite the fact that the items on the two scales were identical, the correlation between them prior to instruction was relatively low, but positive (r = .395).

Self-efficacy for learning in online environments. To date, much of the research concerning self-efficacy in online learning environments has focused specifically on technology use (e.g., DeTure, 2004; Lim, 2001; Miltiadou & Yu, 2000; Puzziferro, 2008; Simmering, Posey, & Piccoli, 2009; Spence & Usher, 2007; Wang, Shannon, & Ross, 2013). The following section will discuss the existing research related to self-efficacy in online learning environments. This will be followed by a description of the Online Learning Self-Efficacy Scale (OLSES) that was developed for use in this dissertation.

Research. As previously stated, much of the research on self-efficacy in online learning environments focuses on technology. For example, DeTure (2004) studied self-efficacy for using synchronous and asynchronous communication using Miltiadou and Yu's (2000) Online Technologies Self-Efficacy Scale (OTSES). She examined the ability of this type of self-efficacy along with cognitive style to predict success in an online course. Neither variable was a significant predictor of online course grades. Puzziferro (2008) also used Miltiadou and Yu's OTSES. She found that scores on the OTSES were not correlated with final grades nor were they related to student satisfaction.

Lim (2001) studied computer self-efficacy in relation to adult online students' satisfaction and intent to enroll in future online courses. She defined computer self-efficacy as "one's beliefs in ability to use computers and to learn new computer skills" (p. 43). A low positive correlation was found between computer self-efficacy and participant's intentions to enroll in online courses in the future ($r = .238$, $p < .001$). A statistically significant relationship between computer self-efficacy and satisfaction was also noted, however the numerical value of that correlation was not given. Simmering, Posey, and Piccoli (2009) also examined computer self-efficacy. They found a positive, but low, correlation between computer self-efficacy and average test grades ($r = .29$, $p < .05$).

Technology use is only one aspect of online learning. A comprehensive scale measuring online learning self-efficacy should also include tasks related to learning skills such as time management and communication. Artino and McCoach's (2008) research included aspects of online learning other than technology use, however their work was specific to self-paced online courses. Nearly all online courses offered at the university where the present research study will occur are paced courses. They typically run the duration of a semester and have weekly assignments. Many paced courses require at least some student-student interaction which is not commonly found in self-paced courses.

A measure recently developed by Shen, Cho, Tsai, and Marra (2013) does measure online learning self-efficacy more comprehensively. Their scale consisted of 35 items rated on an 11-point scale ranging from “cannot do at all” to “highly confident can do” designed to fit into six categories derived following a review of literature: self-efficacy to (a) complete an online course, (b) interact with classmates, (c) interact with an instructor, (d) self-regulate in online learning, (e) handle a course management system, and (f) socialize with classmates. Following an exploratory factor analysis, the scale was reduced to five factors. The self-efficacy to self-regulate in online learning factor was removed and the five additional factors remained. Their scale places the majority of its weight on interactions with three of the five factors directly related to interactions, two with classmates (socially and for academic purposes) and one with instructors. Their sample was comprised completely of students enrolled in online courses, thus lacking diversity.

In the present study, online learning self-efficacy will be defined as learners’ perceptions of their abilities to perform tasks required to be a successful student in a paced online course. Because there was not an existing measure congruent with this definition of online learning self-efficacy that was validated for use with students with and without online learning experiences, a scale was developed and validated for use in this research project: the Online Learning Self-Efficacy Scale.

Online Learning Self-Efficacy Scale. The Online Learning Self-Efficacy Scale (OLSES) was developed as a measure of self-efficacy for succeeding in a formal, paced, online course (see W.A. Zimmerman & Kulikowich, 2013). The OLSES consists of 22 items with each item is rated on a six-point scale. Principal components analyses were used to identify three subscales: learning in the online environment, time management, and technology use. These subscales were deemed appropriate for students at the university with and without formal online learning experiences.

In the validation of the OLSES, subscale scores were correlated with other related and unrelated variables (see W. A. Zimmerman & Kulikowich,2015). The survey was completed by 338 post-secondary students from multiple campuses of a large university system. OLSES subscales were most strongly related to participants' general opinions of online education (learning $r = .651$, $p < .001$; time management $r = .435$, $p < .001$; technology $r = .506$, $p < .001$). Perceived technology skills were most strongly related to the learning ($r = .570$, $p < .001$) and technology ($r = .426$, $p < .001$) subscales. The correlation with the time management subscale was statistically significant, but lower ($r = .230$, $p < .001$). OLSES subscale scores were not statistically related to students' grade point averages (learning $r = .024$, $p = .749$; time management $r = .090$, $p = .228$; technology $r = .039$, $p = .599$).

Subscale scores were compared for participants with and without online learning experiences using a multivariable analysis of variance (MANOVA). While the overall MANOVA was statistically significant [Wilke's Lambda = .904, $F (3, 274) = 9.747$, $p < .001$], the effect size was relatively low (partial $\eta^2 = .096$). On each subscale, participants with online learning experience gave higher ratings than participants without online learning experience. The largest relative differences were on the technology subscale [$F (1, 276) = 28.219$, $p < .001$, partial $\eta^2 = .093$] and the learning subscale [$F (1, 276) = 20.677$, $p < .001$, partial $\eta^2 = .070$]. The smallest difference was observed for the time management subscale [$F (1, 276) = 8.379$, $p = .004$, partial $\eta^2 = .030$]. The statistical significance may in part be due to the relatively large sample sizes, and thus high power. All effect sizes were relatively low. While there are differences between students at this university with and without online learning experiences, the differences were small (W. A. Zimmerman & Kulikowich, 2013, 2015).

The low effect sizes may be due to characteristics of the sample. The university that all participants were recruited from often incorporates online materials into face-to-face courses. For example, the majority of courses use the university's online course management system or a department-run website. Students who have not completed an online course may have had experiences with online learning technologies via their on-campus coursework (W. A. Zimmerman & Kulikowich, 2015).

The OLSES was initially developed specifically for use with students enrolled at this university. Its reliability and validity have been explored with this specific population (W. A. Zimmerman & Kulikowich, 2015). This concludes the discussion of discipline-specific self-efficacies (i.e., statistics self-efficacy and self-efficacy for using educational technology).

Summary of self-efficacy. Self-efficacy is often studied from a social cognitivist perspective. It is most notably associated with the work of Bandura (see 1977a, 1977b, 1978, 1982, 1997, 2007, 2012). Self-efficacy is a construct of interest as it has been shown to be related to behaviors in a wide variety of settings (Bandura, 1982; Schunk, 2012; B. J. Zimmerman, 2000). It is influenced by four often cited sources: modeling/vicarious experiences, personal experiences, psycho-physical status, and feedback from others (Schunk, 2012).

Previous research on statistics self-efficacy has observed moderate, positive relationships with demonstrated knowledge (e.g., Finney & Schraw, 2003; Lane et al., 2004; Zare et al., 2011). Changes in statistics self-efficacy over relatively short periods of time were examined by Finney and Schraw (2003) who did find a significant increase from the beginning to end of a semester and by Lane et al. (2004) who did not find a significant change. However, the measures used by the two teams of researchers differed in their operationalization of the construct. Additionally, the span of time in Lane et al.'s study was less than that of Finney and Schraw's.

Research concerning self-efficacy related to learning in online environments has primarily focused on technology use. Seeing a deficit in the research, Zimmerman and Kulikowich (2015, 2013) developed and validated the OLSES which includes a wider variety of tasks related to learning online. The scale consists of three subscales: learning in the online environment, technology use, and time management. These subscales are consistent across students with and without formal online learning experience.

While the research concerning self-efficacy is vast, a gap in the literature was observed. There is very little research that has examined self-efficacy in conjunction with perceived cognitive load. That is, the relationship between how individuals perceive their abilities to be successful at tasks and their perceptions of their cognitive effort while completing those tasks. The following section describes the research that has been published that has looked at cognitive load and self-efficacy together.

Self-Efficacy and Cognitive Load

Though relatively limited, some research has examined both cognitive load and self-efficacy. Problem solving research, for example, has shown that when a task incorporates appropriate interactive multimedia instruction, learners perceive the task as less cognitively demanding (i.e., lower cognitive load) and their self-efficacy increases. A task that does not do so results in higher cognitive load and lower self-efficacy (Zheng et al., 2009). On a more general level, academic self-efficacy and a measure of working memory capacity have been shown to have a moderate, positive correlation (Vasile et al., 2011). There is a gap in the literature, however, in the examination of self-efficacy for learning in a specific topic area and cognitive load while learning that topic.

The exploration of cognitive load and self-efficacy in an online learning environment is also limited. Online instruction can add extraneous cognitive load to the learning experience, especially for learners who are unfamiliar with the environment. There is an inverse relationship between learners' self-efficacy for completing tasks related to online learning and their self-reported levels of cognitive load. For example, self-efficacy is higher in individuals who report less cognitive demand for tasks required of online learners (McQuaid, 2010). The next step is to expand on this concept to examine the impact of the online environment on learning in a content area, such as statistics.

Summary of Review of Literature

The preceding review summarized the existing literature related to demonstrated topic knowledge, perceived topic knowledge, cognitive load, and self-efficacy. When possible, research concerning students in introductory statistics courses was incorporated. These variables were selected based on their importance in education and their known and unknown relations to one another. One of the primary purposes of formal education is to increase demonstrable topic knowledge. Perceived topic knowledge, cognitive load, and self-efficacy have all been shown to be related to increases in demonstrated knowledge. The interrelationships between these three variables, however, are less clear.

A better understanding of these relationships may shed light on the importance of all three variables in maximizing students' experiences. The impact of instruction that elicits low versus high levels of cognitive load will be compared. This will be done in three studies that will examine learners' self-efficacy, perceived knowledge, perceived cognitive load, and demonstrated knowledge in the topic area of the chi-square test of independence. The following section will provide an overview of the topic of chi-square tests of independence.

Chi-Square Test of Independence

The chi-square test of independence was selected as the topic of focus for this study because it requires only introductory knowledge of statistics and involves relatively basic mathematical operations, yet it is not often taught until later in an introductory or intermediate statistics course.

Thus, introductory statistics students often have little to no prior exposure of the topic but they have the necessary skills needed to learn about the topic.

The chi-square test of independence statistically tests the null hypothesis that two categorical variables are unrelated to one another. For example, one could test the independence of sex (male, female) and eye color (blue, brown, other). This would be a 2x3 chi-square test of independence because there are two levels of the sex variable and three levels of the eye color variable. The null hypothesis would be that there is not a relationship between sex and eye color; that is, they are independent of one another. If the null hypothesis is rejected, then there is statistical evidence that sex and eye color are not independent of one another; in other words, knowing an individual's sex can help you predict their eye color or vice versa (Coladarci et al., 2011; Dodge, 2008).

In order to conduct a chi-square test of independence, there are a number of assumptions that must be met. The variables must be categorical. The sample must be randomly selected so that each member of the population has an equal probability of being selected (Dodge, 2008;see Glass & Hopkins, 1996, p. 226). Each observation may only appear once in the contingency table. All expected values should be at least five (Dodge, 2008).

The test is performed by created a contingency table. For example, in Figure 1, there are 7 males with blue eyes, 9 females with blue eyes, 15 males with brown eyes, and so on. These are the observed values. Row and column totals of observed values are computed as seen in Figure 2. Cell expected values are computed using the equation $e_{ij} = \left(n_{i.} \cdot n_{.j}\right)/n_{..}$ where $n_{i.}$ and $n_{.j}$ are the observed row and column frequencies respectively. This formula is often simplified for introductory statistics students to $\mathrm{E} = [(\mathrm{Row})(\mathrm{Column})]/\mathrm{Total}$. The expected values are often shown in parentheses in the contingency table as seen in Figure 3. These expected and observed values are used to compute the chi-square test statistic: $\chi^2 = \sum_{\mathrm{i}=1}^{\mathrm{c}} \sum_{\mathrm{j}=1}^{\mathrm{r}} \frac{\left(\mathrm{n_{ij}}-\mathrm{e_{ij}}\right)^2}{\mathrm{e_{ij}}}$ where n_{ij} is each cell's observed frequency and e_{ij} is each cell's expected frequency (Dodge, 2008). This formula is often simplified to $\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$ where O is the observed cell frequency and E is the expected cell frequency.

		Eye Color		
		Blue	Brown	Other
Sex	Male	7	15	6
	Female	9	14	8

Figure 1. Chi-Square Test of Independence Example, Contingency Table

		Eye Color			
		Blue	Brown	Other	Row Totals
Sex	Male	7	15	6	28
	Female	9	14	8	31
	Column Totals	16	29	14	59

Figure 2. Chi-Square Test of Independence Example, Observed Values

		Eye Color			
		Blue	Brown	Other	Row Totals
Sex	Male	7 (7.6)	15 (13.8)	6 (6.6)	28
	Female	9 (8.4)	14 (15.2)	8 (7.4)	31
	Column Totals	16	29	14	59

Figure 3. Chi-Square Test of Independence Example, Expected Values

Using the values in Figure 3, $\chi^2 = \frac{(7-7.6)^2}{7.6} + \frac{(15-13.8)^2}{13.8} + \frac{(6-6.6)^2}{6.6} + \frac{(9-8.4)^2}{8.4} + \frac{(14-15.2)^2}{15.2} + \frac{(8-7.4)^2}{7.4} = 0.3925$. The chi-square test statistic for testing the null hypothesis that sex and eye color are independent of one another is 0.3925. This test statistic is compared to a critical value that can be found on the chi-square table. The degrees of freedom are equal to the number of rows minus one, multiplied by the number of columns minus one [$\nu = (R-1)(C-1)$; see Glass & Hopkins, 1996, p. 334]. Thus the degrees of freedom in this example are $\nu = (2-1)(3-1) = 2$. There are two degrees of freedom. The degrees of freedom are used to look up a critical value on a chi-square table. The critical value for a chi-square test with two degrees of freedom at the .05 alpha level is 5.99. In order for the null hypothesis to be rejected, a test statistic of more than 5.99 must be obtained. The test statistic in this example, 0.3925, is not greater than 5.99 therefore there is not sufficient evidence to reject the null hypothesis. There is not evidence that sex and eye color are related to one another.

Conclusion

This chapter reviewed the existing literature in the areas of demonstrated topic knowledge, perceived topic knowledge, cognitive load, and self-efficacy. In terms of the relations between these constructs, a gap was noted in the literature concerning the connection between cognitive load and self-efficacy. In the present dissertation, this gap was explored using the topic of the chi-squared test of independence. This specific statistical test was selected because it requires relatively little knowledge of statistics to learn yet it is not typically taught until at later in introductory or intermediate statistics courses. Additionally, it is a topic whose procedures can be taught in relatively little time. This made it an ideal topic for the present research study.

This research project was conducted in three phases: two pilot studies and the full study. The first pilot study served to develop and validate measures of demonstrated topic knowledge, perceived topic knowledge, and self-efficacy. The second pilot study developed and validated instructional materials designed to elicit low and high levels of cognitive load. The full study used the materials refined in the pilot studies to examine the impact of instructional materials designed to elicit low and high levels of cognitive load on learners’ demonstrable topic knowledge, perceived knowledge, and self-efficacy.

Chapter 3: Pilot Studies

Two pilot studies were conducted with the purpose of developing and validating the materials used in the full study. The first pilot study served to establish the psychometric properties of the scores for the measures of self-efficacy, perceived topic knowledge, and demonstrated topic knowledge. The topic for all measures was the chi-square test of independence. The second pilot study served to validate the instructional materials. These two pilot studies, along with a discussion of their implications for the full study, are outlined in this chapter.

Pilot Study 1: Instrument Development

The purpose of the first pilot study was to develop and establish the psychometric properties of the scores for the instruments designed to measure self-efficacy, perceived topic knowledge, and demonstrated topic knowledge. These three instruments were developed in relation to the topic of the chi-square test of independence. A fourth instrument, the Online Learning Self-Efficacy Scale, was also administered because the majority of the materials in the study were administered online. These instruments were administered to individuals. Participants were classified into two groups: with and without prior experiences related to the chi-square test of independence. These two groups served to represent participants in the full study before and after the presentation of the instructional materials.

The self-efficacy and perceived topic knowledge instruments were each administered twice, both before and after the administration of a demographic survey. Reliability estimates of the self-efficacy ratings were examined using Cronbach's alpha as a measure of internal consistency. Two parallel forms of the demonstrated topic knowledge measure were administered; the same form was not used for both administrations on account of the possibility of participants learning from the initial test taking experience. Demonstrated topic knowledge items were all opened ended. Each item was rated on a three point scale (0 – 2 points each). All items were rated by a second. The relations between scores on the two forms was examined. Cronbach's alpha was also used as a measure of internal consistency.

Evidence for the validity of scores was collected in a number of ways. The composite scores for individuals with and without prior experience with the chi-square test of independence were compared. Correlational analyses were used to provide evidence of convergent and divergent validity.

Instrumentation. Three instruments were developed for use in this research. Those instruments were designed to measure self-efficacy, perceived topic knowledge, and demonstrated topic knowledge. All measures were pertaining to the topic of the chi-square test of independence. Each of the instruments was developed by the researcher for the purposes of this dissertation research.

Chi-square self-efficacy. The instrument for measuring self-efficacy for completing tasks related to the chi-square test of independence was developed to include the major steps required to complete a chi-square test of independence. In designing this scale, the researcher performed a chi-square test of independence and noted each step of the process. The resulting scale consisted of nine items:

1. Identify the assumptions of a χ^2 test of independence
2. Write the appropriate null and alternative hypotheses for a χ^2 test of independence
3. Look up the appropriate critical value on a χ^2 table
4. Calculate the appropriate degrees of freedom for a given χ^2 test of independence
5. Calculate row and column totals
6. Calculate expected cell values
7. Compute the test statistic for a 2x2 χ^2 test of independence
8. Compute the test statistic for a 3x4 χ^2 test of independence
9. Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis

Participants were given the instructions, "Rate your confidence in your current ability to perform the following tasks related to **performing statistics** successfully."

Each item was rated on a nine-point scale. The lower end of the scale contained the anchor "No Confidence." The upper end of the scale contained the anchor "Complete Confidence." Participants were given the following instructions: "A rating of 1 is the lowest; give a rating of 1 to any tasks that you have no confidence that you could complete successfully. A rating of 9 is the highest; give a rating of 9 to any tasks that you have complete confidence that you could complete successfully." These anchors were adapted from the scales used by Bartsch, Case, and Meerman (2012) and by Finney and Schraw (2003).

Perceived topic knowledge. Perceived knowledge of the chi-square test of independence was measured using two items. Participants were asked "How much do you know about the following topics?" Two topics were given: (a) introductory statistics and (b) χ^2 tests of independence. These represented the specific topic that was the focus of the instructional materials, the χ^2 tests of independence, and the broader topic area of introductory statistics.

A nine point scale was used. This scale was selected to be consistent with the self-efficacy scale. Specifically, construction of scales was also similar to the methods used by Tormala and Petty (2007). Additionally, derivatives of the anchors employed by Murphy and Alexander (2004) were used. Participants were given the following directions: "A rating of 1 is the lowest; give a rating of 1 to any topic that you know nothing about. A rating of 9 is the highest; give a rating of 9 to any topic that you know a great deal about."

Demonstrated topic knowledge. Demonstrated knowledge of the topic of the chi-square test of independence was measured using a series of open-ended items. Two sets of demonstrated topic knowledge items were developed: set A and set B. For consistency, the two sets were designed with the intent of being parallel with one another and also with the chi-square self-efficacy measure. Two forms of the demonstrated knowledge measure were constructed: a blue form and a red form. The blue form presented set A first followed by set B. The red form presented set B first followed by set A. The demonstrated knowledge items are presented as Appendix A. Two forms were examined in this pilot study to assure that they were similar enough to be used in the full study as the pre- and post-instruction measures.

For each knowledge set, the first five items were presented on the computer. This allowed participants to type their responses to these items which involved no or very minimal calculations. The last four items required more in-depth calculations. These four items were presented to participants in paper packets.

The demonstrated topic knowledge items were scored by two raters using a three-point scale. This is a method similar to that of Garner and Gillingham (1991). A score of zero was assigned to responses that were missing or completely incorrect. A score of one was given for partially correct responses. And, a score of two was given to correct responses.

Online Learning Self-Efficacy Scale. The Online Learning Self-Efficacy Sale was administered to participants as a check for discriminant validity. By definition, self-efficacy is a task-specific construct (Bandura, 1997). Therefore, online learning self-efficacy and self-efficacy for the chi-square test of independence should not be strongly correlated. There may be a smaller correlation between the two because both constructs are related to learning which could lead to a small amount of shared variance.

The Online Learning Self-Efficacy Scale was previously validated with a diverse sample of students from The Pennsylvania State University (W. A. Zimmerman & Kulikowich, 2015, 2013). It consists of 22 tasks related to online learning. Those tasks were presented to each participant in a random order to prevent any effects possibly related to the seriation of the tasks.

The instructions and anchors used for this scale are similar to those used in the chi-square self-efficacy instrument. Participants were given the instructions, “Rate your confidence in your current ability to perform the following tasks related to **learning online** successfully." Each item was rated on a nine-point scale. The lower end of the scale (i.e., a rating of 1) contained the anchor "No Confidence." The upper end of the scale (i.e., a rating of 9) contained the anchor "Complete Confidence."

Participants. A total of 30 students participated in the study. They were recruited from four different courses related to statistics for the behavioral sciences or educational testing and measurement. All participants were enrolled in masters or doctoral programs in the College of Education. The majority of the sample identified as female (N = 22). The remaining 8 participants identified as male. The number of statistics courses previously completed by participants ranged from 0 to 5. The average number of statistics courses previously completed was 1.933 with a standard deviation of 1.484. Exactly half of participants (N = 15) reported that they had learned about the chi-square test of independence in one of their statistics courses.

Procedures. Prior to data collection, this study was submitted the University's Institutional Review Board and was given the determination of "exempt from IRB initial and ongoing review." Following this determination, instructors of graduate-level courses in the Educational Psychology Program were contacted and asked to participate. Instructors were asked to consider giving extra credit to students in their courses who participated in the study. Students who did not want to participate in the research study were given the option of completing an alternative assignment for an equal amount of extra credit.

Potential participants were sent a copy of the Implied Informed Consent form and a link to sign up for a data collection session. After signing up, and prior to the data collection sessions, participants were emailed a password protected link to the survey. Half of participants were given links to the blue form which presented the demonstrated knowledge sets in the A-B order and half of participants were given linked to red form which presented the demonstrated knowledge sets in the B-A order. Participants were instructed to not open the link in the email until they were in the data collection session and told to do so. The password to access the survey was not given to participants until the beginning of the data collection session.

Data collection sessions were held in computer labs. At the beginning of the session participants were instructed to sign on to their computers and open the link that they had been emailed. Participants were then given the password to access the online instrument. The blue form of the online data collection instrument is attached as Appendix B as screen shots. The first page of the online data collection instrument was the implied informed consent form. The researcher said aloud, “Please read the implied informed consent form. When you are finished, click ‘continue’ and wait for further instructions.”

After all participants had clicked “continue,” the principal investigator read the directions on the second page of the survey out loud and verbatim. Participants then completed the Online Learning Self-Efficacy Scale, chi-square self-efficacy scale, and the perceived topic knowledge measures. After completing these items, participants waited for further instructions from the investigator whom monitored the participants’ progress. After at least 75% of participants had completed all items, the investigator instructed participants to continue to the next section. In each data collection session participants completed each section of the study at the same time. A minimum completion rate of 75% was selected as a cut-off to assure that the majority of participants completed each scale while not taking an excessive amount of time waiting for a small number of participants to complete any section.

Participants were given paper packets with a form ID on the front page. Half of these forms contained blue form items and half contained red form items. The paper packets were matched to the online survey forms. In other words, participants who completed the blue form of the online instrument were given the blue form of the paper packet. Participants who completed the red form of the online instrument were given the red form of the paper packet. The researcher was able to determine the online form type by the Qualtrics URL.

After at least 75% of participants had completed all of the demonstrated topic knowledge items, the researcher directed participants to continue to the next section. This section consisted of demographic items. The demographic questionnaire was presented in the middle of the study as a brief buffer to separate the first and second administrations of the self-efficacy and knowledge scales. The investigator monitored participants' progress. After all participants had completed all items, participants were directed to begin the next section which consisted of the second round of the Online Learning Self-Efficacy Scale, chi-square self-efficacy scale, and the perceived topic knowledge items. After at least 75% of participants completed these items, participants were directed to begin the next section which contained their second set of demonstrated topic knowledge items. Finally, after at least 75% of participants completed the demonstrated topic knowledge items, the researcher instructed participants complete the last part of the survey which requested the information necessary to assign them extra credit in their courses. The data collection session concluded and participants were dismissed.

Results. First, the results from each measure were examined separately. The convergent and divergent validity of scores obtained from each scale was explored using correlational analyses and comparisons of individuals with and without prior related experiences. Finally, the results of all analyses were used to make adjustments to the instruments and procedures for the full study.

Chi-square self-efficacy. The internal consistency of scores obtained from the chi-square self-efficacy measure was evaluated using Cronbach's alpha. The survey was administered two times. For scores obtained during the first administration, Cronbach's alpha was .95. For the scores obtained during the second administration, Cronbach's alpha was .93.

Table 1 contains the descriptive statistics for each item on the instrument and it also contains the correlation coefficients between the first and second administrations as well as effect sizes for the differences between scores obtained at the first and second administrations. Effect sizes were computed by dividing the mean paired difference by the standard deviation for the mean difference for each item.

Table 1: Chi-Square Self-Efficacy Item Scores at Times 1 and 2

	Time 1		Time 2		*r*	*d*
	Mean	SD	Mean	SD		
Identify the assumptions of a χ^2 test of independence	3.38	1.840	2.67	1.807	.790	- 0.533
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	4.00	2.375	2.80	2.058	.674	-0.602
Look up the appropriate critical value on a χ^2 table	4.59	2.666	5.00	2.729	.622	0.213
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	4.34	2.609	3.23	2.161	.650	- 0.488
Calculate row and column totals	5.21	2.808	6.73	2.586	.412	0.562
Calculate expected cell values	3.93	2.604	4.73	2.912	.592	0.359
Compute the test statistic for a 2x2 χ^2 test of independence	3.21	2.161	3.70	2.680	.602	0.257
Compute the test statistic for a 3x4 χ^2 test of independence	3.03	2.129	3.60	2.724	.620	0.265
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	4.86	2.887	4.57	3.014	.711	- 0.074

r represents the Pearson product-moment correlation between scores at time 1 and time 2. *d* is a measure of effect size computed by dividing the mean pairwise difference by the standard deviation for the mean difference for each item

Table 2 contains descriptive statistics for composite scores on the chi-square self-efficacy scale at the first and second administrations. The correlation coefficient between composite scores at the first and second administrations was $r = .739$ (N = 29, $p < .001$).

Table 2: Chi-Square Self-Efficacy Composite Scores at Times 1 and 2

	n	Min	Max	M	SD
Time 1	29	9	70	36.552	19.059
Time 2	30	9	74	37.033	18.531

Perceived Topic Knowledge. Two items were administered to measure perceived knowledge of introductory statistics and the chi-square test of independence. Each item was administered two times.

Descriptive statistics, the correlation coefficient of the scores for the two administrations, and the effect size for pairwise differences are presented in Table 3.

Table 3: Perceived Knowledge Item Scores at Times 1 and 2

	Time 1			Time 2			*r*	*d*
	n	Mean	SD	N	Mean	SD		
Introductory Statistics	29	6.966	1.500	30	6.433	1.851	.888	- 0.554
Chi-square tests of independence	30	3.433	1.977	30	2.933	1.911	.766	- 0.375

Demonstrated topic knowledge. All items were scored by two raters. The first rater was the author. The second rater was a doctoral student studying educational psychology who had completed coursework in statistics and who had knowledge of the chi-square test of independence. Following the initial round of scoring, the raters gave the same scores on 92.407% of items. The correlation coefficient between the scores given by the two raters was computed and the Pearson *r* was .94.

The primary researcher examined all scoring conflicts and made any necessary changes to be consistent with the scoring rubric. The means, standard deviations, correlation coefficients, and effect sizes for pairwise differences between set A and set B are displayed in Table 4. Each item had a minimum score of 0 and a maximum score of 2.

Table 4: Demonstrated Knowledge Item Scores by Set

Item Type	Set A		Set B		*r*	*d*
	M	SD	M	SD		
Identify the assumptions of a χ^2 test of independence	0.300	0.466	0.167	0.379	.683	0.386
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	0.133	0.507	0.333	0.758	.239	0.248
Look up the appropriate critical value on a χ^2 table	1.267	0.980	1.267	0.980	.713	0
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	0.133	0.507	0.133	0.507	1.00	0
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	0.967	0.928	1.200	0.887	.595	0.286
Calculate row and column totals	1.900	0.403	1.967	0.183	- .047	0.148
Calculate expected cell values	1.168	0.986	1.100	0.995	.616	0.077
Compute the test statistic for a 2x2 χ^2 test of independence	0.967	0.890	0.867	0.937	.739	0.151
Compute the test statistic for a 3x4 χ^2 test of independence	0.800	0.761	0.633	0.765	.699	0.281

Internal consistency of scores was assessed. For set A, Cronbach's alpha was .620. For set B, Cronbach's alpha was .635. On both forms, two items had low or negative corrected item-total correlations ("Look up the appropriate critical value on a χ^2 table" and "Calculate row and column totals"). With these two items removed, the resulting scores for the seven item scales had Cronbach's alpha values of .690 and .731, respectively. For the majority of analyses in the first pilot study, responses to all items (i.e., both forms combined) were used. The Cronbach's alpha for the full 18 item scale with no items removed was .810.

On average, scores on set A and set B were similar. There was a strong linear relationship between set A and set B scores ($r = .883$, $p < .001$). A dependent t-test was performed to compare participants' scores on set A and set B and results were not statistically significant [$t(29)=0.111$, $p = .913$, $d = 0.020$].

Set A and set B were completed in different orders by the blue and red groups. Demonstrated knowledge scores by set and by group are presented in Table 5. A repeated measures ANOVA was performed with item set as a within-subjects variable and group as a between subjects variable. There was not a statistically significant set by group interaction [F (1, 28) = 0.103, *p* = .751, η_p^2 = .004] nor was there a statistically significant main effect for set [F (1, 28) = 0.008, *p* = .931, η_p^2 < .001] or for group [F (1, 28) = 0.154, *p* = .698, η_p^2 = .005].

Table 5: Demonstrated Knowledge Set Scores by Group

	Set A			Set B			*r*	*d*
	n	Mean	SD	n	Mean	SD		
Blue Group	14	7.429	3.715	14	7.357	3.565	.865	0.038
Red Group	16	7.813	3.454	16	7.938	3.454	.907	0.086

Online Learning Self-Efficacy Scale (OLSES). Previous validation studies of the OLSES resulted in a three subscale structure: learning, time management, and technology (W. A. Zimmerman & Kulikowich, 2013). The OLSES was administered two times. The scale means, on a scale of 1 to 9, are presented in Table 6 along with the standard deviations and Cronbach's alpha for the first and second administrations. There were not statistically significant differences between the first and second administrations on any of the three subscales [Hotelling's Trace = .162, F (3, 23) = 1.245, *p*= .317, η_p^2 = .140]. Correlation coefficients between the first and second administrations were *r* = .96, .96, and .96 for the three subscales respectively.

Table 6: Online Learning Self-Efficacy Subscales at Times 1 and 2

	Number of Items	Time 1				Time 2			
		n	M	SD	α	n	M	SD	α
Learning	10	30	6.867	1.178	.835	29	6.819	1.410	.922
Time Management	5	30	7.313	1.584	.935	30	7.413	1.506	.947
Technology	7	29	8.000	1.138	.913	28	8.036	1.206	.923

Note: The mean score for each subscale is the average of the items on that subscale; thus, possible subscale means range from 1 to 9.

Correlations. Evidence of convergent and divergent validity was obtained through the examination of the correlation coefficients between variables at both administrations as presented in Table 7. The demonstrated knowledge variable was not separated by administration because the two groups of participants completed the item sets in varying order. For all analyses, unless otherwise noted, demonstrated knowledge scores are composite scores from two item sets combined.

Table 7: Correlation Matrix by Times 1 and 2

		1	2	3	4	5	6	7
1	χ^2 SE		.474*	.890**	.273	-.183	.050	-.144
2	PK: Intro Stats	.499**		.519**	.287	.330	.390*	.321
3	PK: χ^2	.869**	.642**		.182	-.148	-.003	-.076
4	Demonstrated Knowledge	.655**	.333	.378*		-.255	.119	-.142
5	OLSES: Learning	-.141	.287	-.090	-.215		.647**	.809**
6	OLSES: Time Management	.151	.394*	.214	.086	.729**		.462*
7	OLSES: Technology	-.017	.332	.054	-.100	.662**	.408*	

Note: Demonstrated knowledge scores are total from both forms. With that exception, correlations above the diagonal are from the first administration time; correlations below the diagonal are from the second administration time. SE = self-efficacy; PK = Perceived Knowledge; OLSES = Online Learning Self-Efficacy Scale

The strongest correlation coefficients at both the first and second administrations were between the chi-square self-efficacy scale and perceived knowledge of the chi-square test of independence item. These strong positive correlations provide evidence for the convergent validity of the chi-square self-efficacy scale that was developed for use in this research. The newly developed chi-square self-efficacy scale and demonstrated knowledge scales were not strongly correlated with any of the OLSES subscales. These results provide evidence of divergent validity of scores.

Interestingly, the correlation between the chi-square self-efficacy scale and demonstrated knowledge changed from the first to second administrations (r = .273 to r = .655). Note, however, that the demonstrated knowledge scale in these correlations is actually the combined first and second administrations. This change in correlations between self-efficacy and demonstrated knowledge may be a sign of improved self-efficacy calibration. Calibration is the relationship between perceived ability and observable ability (Glenberg, Sanocki, Epstein, & Morris, 1987). This may be of interest in future studies.

Comparisons by prior experience. As part of the demographic questionnaire, participants were asked if they had learned about the chi-square test of independence in a statistics course. Demonstrated knowledge scores were compared for participants relative to their prior experience with the test. The demonstrated knowledge items completed by participants in the blue and red groups were different for the two administrations, thus only total demonstrated knowledge scores were analyzed. An independent t-test was used to compare the demonstrated knowledge total scores of individuals who had not learned about the chi-square test of independence in a previous statistics course (n = 15, *M* = 13.8, *SD* = 6.879) to individuals who had learned about the test previously (n = 15, *M* = 16.8, *SD* = 6.201). Within groups variation was relatively high. The results were not statistically significant [*t* (28) = 1.255, *p* (1-tailed) = .110]. However, there was a medium effect size (*d* = 0.459) which indicates practical significant. Observed power (.338) for the directional test was low. Therefore, these results may have been effected by the relatively small sample size.

The chi-square self-efficacy ratings of participants with and without prior experiences with the chi-square test of independence at the first and second administrations of the scale were also compared. Descriptive statistics are presented in Table 8. A repeated measures analysis of variance was used. Prior experience served as a between-subjects independent variable and administration served as a repeated within-subjects independent variable. Composite scores on the nine self-efficacy items served as the dependent variable. There was not a statistically significant interaction between prior experience and administration [Wilks' Lambda = .954, *F* (1, 28) = 1.345, *p* = .256, $\eta_p^2 = .046$], nor was there a main effect for administration [Wilks' Lambda = .991, *F* (1, 28) = 0.261, *p* = .614, $\eta_p^2 = .009$], thus providing evidence of stability. There was a significant main effect for prior experience [*F* (1, 28) = 17.890, *p* < .001, $\eta_p^2 = .390$]. Individuals with prior experience with the chi-square test of independence gave significantly higher ratings on the self-efficacy items than individuals without prior experience, and with a large effect size.

Table 8: Self-efficacy by administration and prior experience

	n	Administration 1		Administration 2	
		M	SD	M	SD
With Prior Experience	15	47.933	13.424	46.200	15.289
Without Prior Experience	15	23.408	16.502	27.967	17.250

Pilot 1 Summary. The purpose of this pilot study was to assess the psychometric properties of the scores of the instruments designed for the full study. Both reliability and validity of scores were evaluated. In this section, the results of these analyses are summarized and implications for the full study are discussed.

Reliability. Cronbach's alpha was used as a measure of consistency for the chi-square self-efficacy and demonstrated topic knowledge scales. Internal consistency of chi-square self-efficacy scale scores was very high for both the first and second administrations (α = .954 and .933, respectively). Given these results, no changes were deemed necessary to the chi-square self-efficacy items.

For the demonstrated knowledge forms, each with 9 items, alpha was relatively low at .620 for set A and .635 for set B. When responses for the two sets were combined, 18 items total, Cronbach's alpha increased to .810; this increase is likely due in-part to the increase in the number of items and the repetition in the items. Consistency of participants' scores across forms was also evaluated. There were no statistically significant differences for scores obtained from the two sets of items. There was also not a significant effect of the order presentation of the two item sets in that the group that received set A followed by set B (i.e., the blue group) did not differ from the group that received set B followed by set A (i.e., the red group). The correlation coefficient between scores on the two item sets was very strong (r = .883), however this is not sufficient evidence to state that set A and set B are parallel. In the full study, scores on the pre-instruction form will be treated as a covariate as opposed to serving as a repeated measure and they will not be treated as parallel forms.

Inter-rater consistency was also evaluated for the demonstrated knowledge items. All demonstrated knowledge responses were scored by two raters. The two raters assigned the same scores on 92.407% of item responses. The correlation coefficient between the scores given by the two raters was high (r = .944). The same method of scoring responses will be employed in the full study.

Validity. Evidence for convergent and divergent validity was collected. Correlation coefficients between the chi-square self-efficacy scale, perceived knowledge items, demonstrated knowledge scale, and OLSES subscales were examined. Chi-square self-efficacy scale scores were strongly related to ratings of perceived knowledge of the chi-square test of independence (r = .890, .869, for the first and second administrations of each, respectively). The correlation coefficients between chi-square self-efficacy scale scores and perceived knowledge of introductory statistics were positive (r = .474, .499, for the first and second administrations of each, respectively) but weaker than that of the correlation coefficients with the more specific perceived knowledge of the chi-square test of independence.

Correlation coefficients with the OLSES subscale scores provide evidence of divergent validity. As seen in Table 7, the only significant correlation coefficients between statistics-related measures and OLSES subscales were between the perceived knowledge of introductory statistics item and the time management subscale of the OLSES (r = .390, .394, for the first and second administrations of each, respectively). The low correlations between chi-square self-efficacy scale ratings and OLSES subscale scores support the premise that self-efficacy is task specific as there was very little shared variance between the measures and provides evidence for divergent validity.

Also, individuals with and without prior experience with the chi-square test of independence were compared in terms of their demonstrated knowledge scores and chi-square self-efficacy scores. Individuals who reported having learned about the chi-square test of independence in a statistics course had higher demonstrated knowledge scores than individuals who did not report having such an experience; however, the difference was not statistically significant. There was a statistically significant difference in terms of chi-square self-efficacy. Participants who reported having prior experience with the topic gave higher self-efficacy ratings than participants who reported not having related prior experiences.

Changes for full study. As a result of the pilot study, no changes were made to the perceived knowledge items or the OLSES. Changes were made to the demonstrated knowledge scale. The chi-square self-efficacy scale was also adjusted to maintain its alignment with the demonstrated knowledge items. Changes were made to the demographic questionnaire.

On the demonstrated knowledge scale, the 3x4 chi-square computation was removed. The full study will be limited to 2x2 chi-square tests to reduce the amount of time needed to complete the measure and to reduce the depth of the instruction that will be necessary. For consistency, the chi-square self-efficacy scale was adjusted to reflect this change in that the item concerning computing the test statistic for a 3x4 chi-square test was removed.

The demographic questions used in the present study were designed for use with a graduate student population. The full study will sample primarily from undergraduate-level courses. The item “Do you have any experience with online education other than for-credit online courses?” was removed. Questions related to participants prior experiences and related abilities were added including items concerning SAT and ACT quantitative scores and mathematics courses completed in high school and college.

In addition to changes in the items administered, the administration of the measures was changed following this study. In this pilot study, participants completed each section as a cohort in that they began each section of the survey together then waited for others to finish. Participants who completed a section quicker than the others in their session sat quietly and waited for others to finish.

In future studies participants will not all complete the research study in unison. Instead, they will begin the study together then complete the entire study at their own pace.

Pilot Study 2: Instructional Materials

The second pilot study served to validate instructional materials designed to elicit low and high levels of cognitive load. The study used methods similar to those of Sweller and Cooper (1985) and Paas (1992). The topic of the materials was the chi-square test of independence. Like Sweller and Cooper's study, this pilot study compared conventional problem solving and worked examples. The measurement of perceived cognitive load was similar to the methods used by Paas. Van Gog et al.'s (2012) suggestions for timing were also taken into account. A variation of Paas' perceived mental effort scale was administered to participants after watching an instructional video and after each of three practice problem set that they studied or solved.

In the second pilot study, a video introducing the topic of the chi-square test of independence was viewed by participants. Two sets of practice materials were created: a conventional problem solving set and a worked example set. These materials were administered to a sample of undergraduate introductory statistics students.

Participants. Participants were recruited from two undergraduate introductory statistics courses. Instructors were contacted and asked to give their students extra credit for participation in the study. A total of nine participants completed the study. They were randomly assigned to either the conventional problem solving or worked example condition with four participants assigned to the former and five assigned to the latter condition.

Materials. The materials included an instructional video consisting of an introduction to performing the chi-square test of independence, practice problems for each the conventional problem solving and worked example groups, a measure of perceived cognitive load, and a demographics questionnaire. Here, each of these materials will be described in greater detail. Note that for the conventional problem solving and worked example group, the only difference in the materials presented is in the first of the three problem sets. All other materials were identical for the two groups.

***Instructional video*.** The instructional video introduced the topic of the chi-square test of independence and provided one worked example of a 2x2 test. The researcher created a PowerPoint presentation with a script. The PowerPoint slides are attached as Appendix C. The script is attached as Appendix D. The presentation included a brief description of the chi-square test of independence including the assumptions of the test and its hypotheses. The calculation of degrees of freedom was described as well as how to use the degrees of freedom to look up a critical value on the chi-square table. Finally, an example of the calculation of the test statistic was also given.

The PowerPoint and script were used to make a video. QuickTime and iMovie softwares were used in the production. The video consisted of the PowerPoint slides visually and one of the secondary investigators reading the script out loud. Its duration is 11 minutes and 1 second. The resulting video may be played in iTunes on both Macintosh and Windows computers. For this study, the video was embedded in the online Qualtrics instrument.

Practice problems. Two forms of the practice problems were created. The problems on the two forms were the same however the methods utilized were different. Form A consisted of three conventional problems while form B consisted of one worked example and two conventional problems.

Both form A and form B consisted of three practice problems. All problems were in the form of 2x2 chi-square tests of independence. Form A, which was received by the conventional problem solving group, consisted entirely of uncompleted problems (see Appendix E). Form B, which was received by the worked example group, began with one worked example followed by two uncompleted problems (see Appendix F).

For each of the conventional problem sets, participants were presented with a brief scenario and a 2x2 contingency table. They were asked three questions. First, they were asked for the correct critical value via an open-ended item. The second question asked for the test statistic for the given scenario, again this was an open-ended item. Finally, they were asked to identify the appropriate conclusion. The format of this item was multiple choice. If a participant submitted an incorrect response, the participants was told that the response was incorrect and was given two additional opportunities to fix the answer before being given the solution. For the second question concerning the test statistic, a 0.10 margin of error was permitted to allow for any minor differences due to rounding.

Perceived cognitive load. As per the suggestions of Van Gog et al. (2012), perceived cognitive load was measured at four points using a method similar to that of Paas (1992). Participants were asked to rate their mental effort after watching the instructional video which served as a baseline. Participants were also asked to rate their mental effort after each of the three practice problems, regardless of whether they received a worked example or conventional problem. A nine-point scale was utilized with the anchors of *1 = very minimal mental effort* to *9 = highest possible mental effort*.

Procedures. Prior to data collection, this study was submitted the University's Institutional Review Board and was given the determination of "exempt from IRB initial and ongoing review." Data collection occurred in computer labs. Participants signed up for data collection sessions via a SignUpGenius.com online signup sheet. They were asked to bring a pencil, calculator, and headphones. The day of their data collection session, participants were emailed a link to the Qualtrics survey which was password protected. For each session, approximately half of the participants were assigned to the conventional problem solving group and half were assigned to the worked example group using a random number generator.

To begin the session, each participant was seated at a computer. Participants were instructed to sign into the computers using their university access accounts. Then, participants were instructed to open the link to the Qualtrics survey that they had been emailed. Participants were given the password needed to access the survey. Each participant was given a sheet of formulas and the chi-square distribution. The researcher instructed the participants to read the informed consent form and press continue. Then, the researcher informed the participants that they would complete the remainder of the session at their own pace.

The instructional interventions for the two groups are presented in Figure **4**. Participants first viewed the instructional video on their own computers listening via headphones. Following the video, participants rated their level of perceived cognitive load while watching the video. Then, participants were presented with the appropriate practice problems. After each of the three problem sets, participants were asked to rate their perceived level of cognitive load.

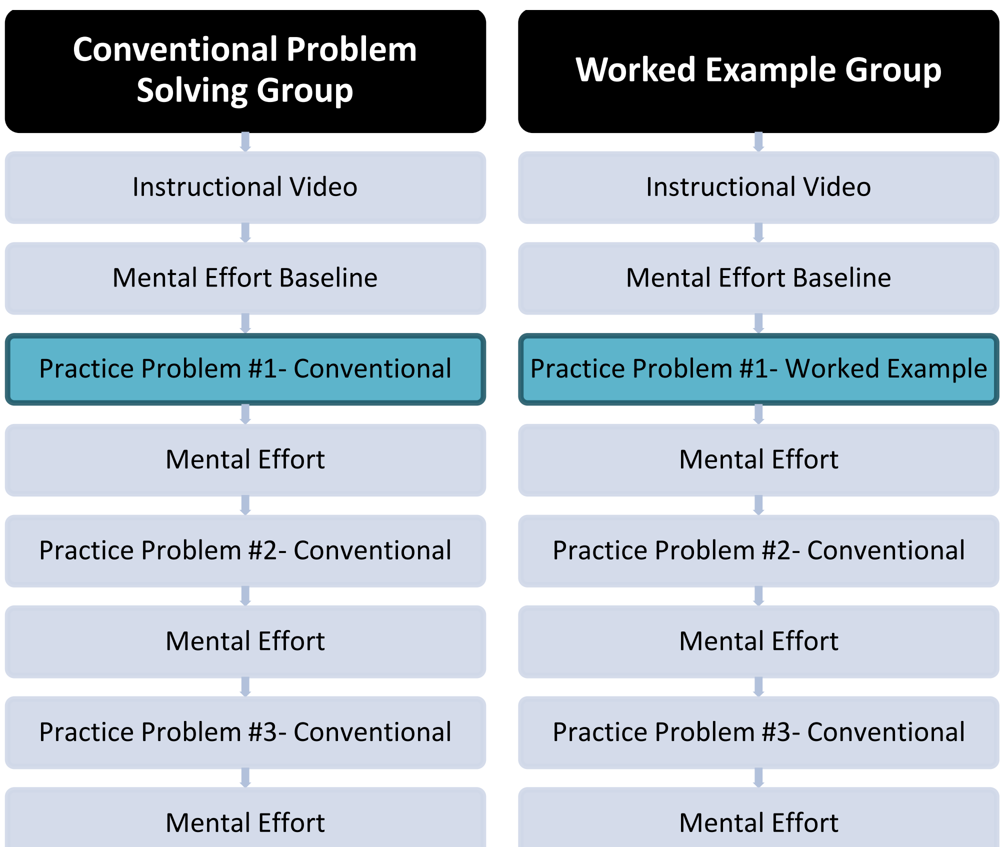


Figure 4: Design of Instructional Intervention

Following the instructional intervention, participants were asked a series of demographic questions similar to those used in the first pilot study. They were also asked for the identifying information necessary for assigning extra credit.

Results. The perceived cognitive load ratings collected following the viewing of the instructional video and the solving of the practice problems given by participants in the worked example and conventional problem solving conditions were compared descriptively. Inferential statistics were not employed due to the very small sample size. A summary of the perceived cognitive load ratings following the viewing of the video and for the average following each of the three practice problem sets for each of the two groups is presented in Table 9. Ratings given following the instructional video were more than half a standard deviation higher for the worked example group than for the conventional problem solving group. For the conventional problem solving group, an increase in perceived cognitive load of 0.555 standard deviations was observed from the baseline of watching the instructional video to the completion of the practice problems. For the worked example group, no change in perceived cognitive load was observed. This supports the premise that instructional materials using only conventional problem solving methods may lead to an increase in cognitive load.

Table 9: Perceived Cognitive Load Ratings without Adjustments

		Following Instructional Video		Average Following Practice Problem Sets	
Condition	n	M	SD	M	SD
Conventional	5	3.800	2.049	5.400	2.229
Worked Example	4	5.000	2.160	5.000	1.963

The average perceived cognitive load ratings following the practice problem sets were adjusted to take into account differences at the baseline. Perceived cognitive load ratings following the instructional video were used as a covariate (on the basis of a baseline, perceived cognitive load of 4.333 following the instructional video). With this adjustment, the mean perceived cognitive load rating for the conventional problem solving group was 5.592 (SD = 2.187) and for the worked example group was 4.760 (SD = 2.200). Using a pooled standard deviation computed using the adjusted values ($s_p = 2.193$), a medium effect size was computed for the differences in perceived cognitive load for the conventional problem solving and worked example groups ($d = 0.379$).

Pilot 2 Summary. The results of the second pilot study provide evidence for the validity of developed practice problem sets for eliciting differing levels of perceived cognitive load. There is evidence that the worked example practice problem sets elicit lower ratings of perceived cognitive load when compared to the conventional practice problem sets. Participants in the conventional problem solving group saw an increase in perceived cognitive load ratings while the worked example group did not see any change in perceived cognitive load ratings. When controlling for baseline differences, there was a medium effect size for the differences in perceived cognitive load between the two groups ($d = 0.379$) with the conventional problem solving group giving, on average, higher ratings of perceived cognitive load than the worked example group. No changes will be made to the instructional video or the practice problems sets for the full study.

Chapter 4: Full Study

Research suggests that instructional materials can impact cognitive load (Plass et al., 2010b; Sweller, 2010b). In the following study, cognitive load was manipulated through the use of different instructional materials. Methods similar to those used by Sweller and Cooper (1985) and Paas (1992) were employed. Suggestions from van Gog et al.'s (2012) research concerning the measurement of cognitive load were also taken into account.

As stated in Chapter 1, two research questions were addressed in this study. First, to what extent do instructional materials designed to elicit low and high levels of cognitive load impact self-efficacy, perceived knowledge, and demonstrated knowledge? Second, to what extent do instructional materials designed to elicit low and high levels of cognitive load impact the relations between self-efficacy and demonstrated knowledge?

The instruments developed in the first pilot study were administered to participants before and after the instructional materials. The instructional video used in the second pilot study along with the practice problem sets validated in that study served as the instructional intervention. Participants were randomly assigned to the worked example or conventional problem solving condition. Following the instructional video, the worked example group was presented with one fully worked example to study followed by two conventional practice problems to work through on their own while the conventional problem solving group received all three practice problems in the conventional problem solving format.

Methods

Participants. Participants were recruited from an introductory statistics course offered in the Department of Statistics. The instructor of the course agreed to award students extra credit in exchange for their participation in the research study. Students who were unable to participate in the study or who did not want to participate in the study were given the option of completing an alternative assignment for an equivalent amount of extra credit. There were 312 students enrolled in the course and 103 students participated in the data collection. One participant was observed looking up answers online during the pre-test therefore that participant's data were removed for all analyses resulting in a maximum valid sample size of 102.

Participants ranged in age from 18 years to 24 years with a mean of 19.465 and standard deviation of 1.162 years (n=101). Of the 102 participants, 48 identified as female and 54 identified as male. The majority of participants were sophomores (n=76), followed by freshman (n=10), juniors (n=9), and seniors (n=6). The majority had not previously learned about the chi-square test of independence in a statistics course (88 out of 101 valid responses). Participants represented a variety of majors. The most highly represented majors reported included Business, Undecided, Psychology, Biology, and Kinesiology.

Materials. All materials used in this study were validated in the first two pilot studies. These materials included measures of chi-square self-efficacy, perceived knowledge, and demonstrated knowledge of the chi-square test of independence which were developed in the first pilot study. The chi-square self-efficacy scale and the demonstrated knowledge scales were changed as a result of the first pilot study. Specifically, the items concerning 3x4 chi-square tests were removed. Removal of these items was done to reduce the time needed to complete the demonstrated knowledge scale and to simplify the necessary instructional materials. Thus, each scale was reduced to 8 items. The OLSES was also administered at the beginning of the study.

The instructional video and practice problem sets developed in the second pilot study served as the instructional intervention. The nine-point perceived cognitive load scale was also consistent with the second pilot study and no changes were made. A demographics questionnaire was administrated at the end of the study.

Procedures. Prior to data collection, this study was submitted the University's Institutional Review Board and was given the determination of "exempt from IRB initial and ongoing review." Data collection occurred in computer labs. Participants signed up for data collection sessions via a SignUpGenius.com online signup sheet. They were asked to bring a pencil, calculator, and headphones. The day before or the day of their data collection session, participants were emailed a link to the online Qualtrics survey with instructions to not open the link until they were in the computer lab and instructed to do so by the researcher. The link was password protected. For each session, approximately half of the participants were assigned to the conventional problem solving group and half were assigned to the worked example group using a random number generator.

Data collection sessions were overseen by the primary researcher (i.e., the author). For sessions with more than 20 participants, a second investigator assisted with research procedures as approved by the University's Institutional Review Board. To begin the session, each participant was seated at a computer. Participants were instructed to sign into the computers using their university access accounts and to open the link to the Qualtrics survey that they had been emailed in Firefox. Either the primary researcher or a secondary investigator entered the password into the Qualtrics link which presented participants with the implied informed consent form. Each participant was given a sheet with formulas on one side and the chi-square distribution on the opposite side. Each participant was given a packet of the demonstrated knowledge items as well as a sheet of blank paper to use when completing the online practice problems.

After all participants had agreed to the online implied informed consent form, the primary researcher or a secondary investigator read the instructions out loud to participants. Then, participants worked through the study materials at their own paces. The materials were presented in the following sequence:

1. OLSES, chi-square self-efficacy scale, perceived knowledge
2. Demonstrated knowledge
3. Instructional video

4. Practice problem sets (worked example or conventional), mental effort item after each set
5. Chi-square self-efficacy scale, perceived knowledge
6. Demonstrated knowledge
7. Demographic questionnaire
8. Identifying information for extra credit purposes

Participants were permitted to leave after they had submitted the last page of the online survey. The researchers collected the demonstrated knowledge forms. Participants could choose to keep their scrap paper and formula sheet or recycle them.

Descriptive Statistics

Online Learning Self-Efficacy Scale. Complete data were available from 99 participants. The following analyses include only the data from those 99 participants. The means and standard deviations for these items are presented in Table 10 in order from highest to lowest rated. Each of these items was rated on a nine-point scale where 1 was anchored with “no confidence” and 9 was anchored with “complete confidence.”

Cronbach’s alpha was used as a measure of internal consistency. An alpha of .944 was computed for this 24 item scale. In addition to full scale statistics, subscale descriptive statistics and internal consistency were also evaluated. Table 11 presents the means, standard deviations, and Cronbach’s alpha reliability coefficients for each of the three subscales: learning, time management, and technology. Each of these is on the scale of 1 to 9.

Table 10: Online Learning Self-Efficacy Scale Descriptive Statistics

Item	M	SD
Find the course syllabus online	8.27	1.713
Submit assignments to an online dropbox	8.13	1.718
Complete all assignments on time	7.78	1.588
Navigate the online grade book	7.63	1.877
Search the Internet to find the answer to a course-related question	7.57	1.779
Search the online course materials	7.51	1.769
Learn without being in the same room as other students	7.40	1.823
Navigate online course materials efficiently	7.33	1.818
Develop and follow a plan for completing all required work on time	7.27	1.634
Communicate effectively with my instructor via email	7.15	2.017
Complete an online statistics course with a grade of A or B	7.14	1.818
Learn to use a new type of technology efficiently	7.11	1.571
Communicate effectively with technical support via email, telephone, or live online chat	7.08	2.029
Learn from videos	7.04	1.812
Communicate using asynchronous technologies (discussion boards, email, etc.)	7.02	1.938
Meet deadlines with very few reminders	7.00	1.868
Use synchronous technology to communicate with others (such as Skype)	7.00	2.090
When a problem arises, promptly ask questions in the appropriate forum (email, discussion board, etc.)	7.00	1.744
Overcome technical difficulties on my own	6.86	1.744
Learn without being in the same room as the instructor	6.68	1.937
Manage time effectively	6.55	1.842
Complete a group project entirely online	6.49	2.002
Focus on schoolwork when faced with distractions	6.04	1.932
Use the library's online resources efficiently	6.03	1.997

Table 11: Online Learning Self-Efficacy Subscale Statistics

	Number of Items	M	SD	Cronbach's α
Learning	10	6.868	1.321	.883
Time Management	5	6.935	1.630	.816
Technology	7	7.661	1.412	.893

Chi-square self-efficacy. Means and standard deviations for pre- and post-instruction administrations for each of the chi-square self-efficacy items for both conditions combined are presented in Table 12 (for differences by condition see Table 20). This table also includes the correlation coefficients between the pre- and post-instruction administrations and the effect size (Cohen's *d*) for the difference between the two administrations. All effect sizes are positive signifying an average increase in ratings on every chi-square self-efficacy item from the first to the second administration.

Table 12: Chi-Square Self-Efficacy Item Statistics

Item		Pre- Instruction		Post- Instruction		*r*	*d*
	n	M	SD	M	SD		
Identify the assumptions of a χ^2 test of independence	101	3.40	2.241	5.98	2.306	.166	0.865
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	100	3.20	2.025	5.84	2.273	.742	0.882
Look up the appropriate critical value on a χ^2 table	100	3.75	2.293	6.86	2.370	.819	0.932
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	101	3.05	2.109	6.80	2.375	.812	1.167
Calculate row and column totals	101	5.25	2.896	8.03	1.830	.025	0.908
Calculate expected cell values	100	4.37	2.529	7.06	2.343	.020	0.890
Compute the test statistic for a 2x2 χ^2 test of independence	100	3.00	2.074	6.27	2.550	.473	1.032
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	101	3.37	2.212	6.03	2.330	.849	0.836

Cronbach's alpha was computed as a measure of internal consistency for the chi-square self-efficacy scale scores. For this analysis, all participants with complete data were represented in the sample. Analyses split by condition are presented later, when research question 1 is addressed. For the pre- instruction time point, an alpha coefficient of .913 was computed (valid n = 102). For the post-instruction time point, an alpha coefficient of .921 was computed (valid n = 96).

Pre- and post- instruction average scores were computed. For individuals with complete data and those missing data for one or two items, these scores were computed using the average of the available responses. For the pre-instruction, on a scale of 1 to 9, the mean was 3.670 (SD=1.835, valid n = 102). For the post- instruction, on a scale of 1 to 9, the mean was 6.611 (SD = 1.863, valid n = 101).

The correlation coefficient between pre- and post- instruction composite scores was r = .090 (p = .371, n = 101).

Demonstrated topic knowledge. A subset of ten percent of demonstrated knowledge forms were scored by two raters to establish agreement. The primary scorer was the author and the second rater was the doctoral student who also served as a rater for the first pilot study. Following the initial review, the two raters agreed on the scores to be assigned on 90% of the items. There was a one-point different for 8.5% of items and a two-point difference on the remaining 1.5% of items scored. Following further review, it was noted that of the 10% of items on which the raters disagreed, approximately half were on the questions concerning assumptions of the chi-square test of independence. After taking into account this discrepancy, the two raters agreed on 95% of items that they both scored.

Means and standard deviations for pre- and post-instruction administrations for each of the demonstrated knowledge items for both conditions combined are presented in Table 13. This table also includes the correlation coefficients between the pre- and post-instruction administrations and the effect size (Cohen's *d*) for the difference between the two administrations. All effect sizes are positive signifying an average increase in scores on every demonstrated knowledge item. Correlation coefficients between chi-square self-efficacy and demonstrated knowledge composite scores at the pre- and post-test points are presented in Table 14.

Table 13: Demonstrated Knowledge Item Statistics

Item Type	Pre-Instruction (n = 102)		Post-Instruction (n = 100)		*r*	*d*
	M	SD	M	SD		
Identify the assumptions of a χ^2 test of independence	0.020	0.139	0.850	0.914	.024	0.705
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	0.098	0.411	0.780	0.970	.128	0.681
Look up the appropriate critical value on a χ^2 table	0.706	0.960	1.520	0.858	.216	0.714
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	0.412	0.813	1.320	0.952	.318	0.873
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	0.098	0.330	1.240	0.806	.173	1.398
Calculate row and column totals	1.961	0.241	1.970	0.223	.725	0.058
Calculate expected cell values	0.461	0.817	1.530	0.797	.279	1.068
Compute the test statistic for a 2x2 χ^2 test of independence	0.471	0.805	1.590	0.740	.231	1.162

Table 14: Correlation coefficients for Self-Efficacy and Demonstrated Knowledge

		1	2	3	4
1. Pre-Instruction Chi-Square Self-Efficacy	*r*	---			
	p				
	n	102			
2. Pre- Instruction Demonstrated Knowledge	*r*	-.037	---		
	p	.715			
	n	102	102		
3. Post- Instruction Chi-Square Self-Efficacy	*r*	.090	.147	---	
	p	.371	.143		
	n	101	101	101	
4. Post- Instruction Demonstrated Knowledge	*r*	-.118	.338**	.591**	---
	p	.242	.001	< .001	
	n	100	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

Perceived knowledge. The descriptive statistics for responses to the two perceived knowledge items are presented in Table 15 and Table 16. Each item was rated on a scale of 1 to 9. Kendall's tau was used as a measure of this relation due to the ordinal scales of the variables. The correlation coefficients between the pre- and post-instruction administration responses were τ = .263 (*p* = .001, n = 94) for perceived knowledge of introductory statistics and τ = .032 (*p* = .688, n = 100) for perceived knowledge of the chi-square test of independence. For both items there was an increase in perceived knowledge from the pre to post-instruction administrations.

Table 15: Perceived Knowledge of Introductory Statistics

	n	Q1	Median	Q3	M	SD
Pre-Instruction	97	5	5	7	5.51	1.589
Post-Instruction	98	5	6	7	5.73	1.517

Table 16: Perceived Knowledge of Chi-Square Test of Independence

	n	Q1	Median	Q3	M	SD
Pre-Instruction	102	1	2	3	2.46	1.663
Post-Instruction	100	3	5	7	5.07	2.061

Efficacy of Intervention

The efficacy of the intervention was evaluated by examining perceived cognitive load ratings of participants in the conventional problem solving and worked example groups. Perceived cognitive load was rated on a scale of 1 to 9 after watching the instructional video and after completing each of the three practice problem sets. The descriptive statistics for these perceived cognitive load ratings are presented in Table 17 for both the conventional problem solving and worked example groups.

Table 17: Perceived Cognitive Load Ratings by Group and Practice Set

Group		Video	Practice Set 1	Practice Set 2	Practice Set 3
Conventional	M	5.66	6.54	6.56	6.53
	N	50	50	50	49
	SD	1.955	2.002	1.991	2.180
Worked Example	M	5.31	5.20	6.04	6.18
	N	51	51	51	51
	SD	1.738	1.800	1.939	1.797
Total	M	5.49	5.86	6.30	6.35
	N	101	101	101	100
	SD	1.847	2.010	1.973	1.992

Differences in the perceived cognitive load ratings given while problem solving (i.e., total of the three practice sets' ratings) while controlling for baseline perceived cognitive load ratings (i.e., ratings following the video) for participants in the conventional and worked examples groups were compared using an analysis of covariance. Group served as the between-subjects independent variable. Perceived cognitive load ratings given following the video served as a covariate. The dependent variable was the total perceived cognitive load rating given following the three practice sets.

After adjusting for ratings given following the video (adjusted to a perceived cognitive load rating following the video of 5.51), the mean perceived cognitive load rating while problem solving for the conventional group was 19.506 (SD = 4.487, n=49) and for the worked example group was 17.710 (SD = 4.485, n=51). There was a statistically different main effect for group. The full results of this analysis of covariance are presented in Table 18. On average, participants in the conventional problem solving group gave ratings of perceived cognitive load while problem solving that were 0.400 standard deviations higher than the worked example group when controlling for the baseline ratings given following the video; this is a moderate effect size.

Table 18: Effect of Group on Total Problem Solving Cognitive Load Controlling for Video Ratings

Source	Type III SS	df	MS	F	p	η_p^2	Observed Power
Corrected Model	908.428	2	454.214	22.713	< .001	.319	1.000
Intercept	1028.757	1	1028.757	51.444	< .001	.347	1.000
Video Rating (Covariate)	763.938	1	763.938	38.202	< .001	.283	1.000
Group	79.644	1	79.644	3.983	.049	.039	.506
Error	1939.762	97	19.998				
Total	37407.000	100					
Corrected Total	2848.190	99					

R^2 = .319, Adjusted R^2 = .305

Research Question 1

To what extent do instructional materials designed to elicit low and high levels of cognitive load impact self-efficacy, perceived knowledge, and demonstrated knowledge?

In this section, the differences between the conventional problem solving and worked example will be examined for each variable separately. The chi-square self-efficacy and demonstrated topic knowledge variables met the normality requirements necessary for performing parametric statistics. The perceived knowledge items however did not. Thus, nonparametric analyses were used for the analyses of the perceived knowledge items.

Chi-square self-efficacy. Descriptive statistics for average chi-square self-efficacy scale ratings are presented in Table 19 with a corresponding plot in Figure 5. Descriptive statistics of the chi-square self-efficacy scale items were computed for pre- and post-instruction time points for the conventional problem solving and worked example groups. These values are presented in Table 20.

Table 19: Chi-Square Self-Efficacy Average Composite Scores by Condition

		Pre-Instruction		Post-Instruction	
Group	n	M	SD	M	SD
Conventional	49	3.347	1.861	6.488	2.014
Worked Example	52	3.978	1.791	6.726	1.721
Total	101	3.672	1.844	6.611	1.863

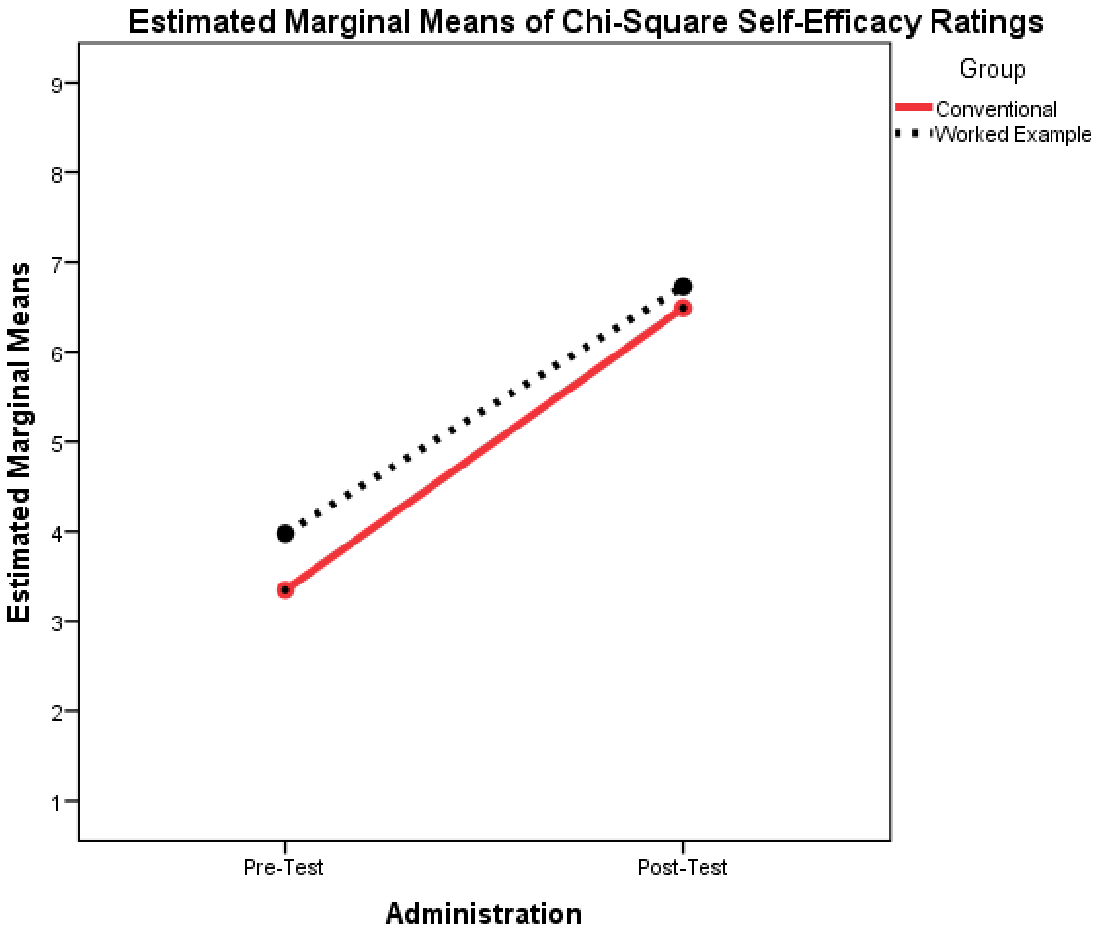


Figure 5: Plot of Margin Means of Chi-Square Self-Efficacy Ratings by Group

Table 20: Self-Efficacy Item Statistics by Condition and Administration

	Conventional Problem Solving				Worked Example			
	Pre		Post		Pre		Post	
Item	M	SD	M	SD	M	SD	M	SD
Identify the assumptions of a χ^2 test of independence	3.16	2.244	5.69	2.356	3.65	2.222	6.25	2.248
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	2.88	1.913	5.71	2.333	3.56	2.071	5.96	2.231
Look up the appropriate critical value on a χ^2 table	3.48	2.279	6.90	2.520	3.98	2.271	6.83	2.247
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	2.72	1.949	6.53	2.534	3.37	2.205	7.06	2.209
Calculate row and column totals	4.82	2.812	8.04	1.925	5.62	2.938	8.02	1.754
Calculate expected cell values	3.92	2.465	7.00	2.560	4.71	2.554	7.12	2.148
Compute the test statistic for a 2x2 χ^2 test of independence	2.92	2.137	5.98	2.709	3.13	2.029	6.54	2.388
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	2.90	1.982	6.02	2.487	3.81	2.327	6.04	2.196

To compare the mean chi-square self-efficacy ratings given by the conventional and worked example groups at the pre- and post-instruction points, a repeated measures analysis of variance was performed. The assumption of equal covariance matrices was met [Box's M = 1.345, F (3, 1994024.556) = 0.438, *p* = .725] as was the assumption of homogeneity of variances which was assessed via Levene's test of equality of error variances [pre-instruction F (1, 99) = 0.134, *p* = .715; post- instruction F (1, 99) = 0.820, *p* = .367]. There was not a statistically significant group by administration interaction [Greenhouse-Geisser = 1.956, F (1, 99) =0.623, *p* = .432, $\eta_p^2 = .006$]. There was a statistically significant main effect for administration [Greenhouse-Geisser = 437.452, F (1, 99) = 139.379, *p* <.001, $\eta_p^2 = .585$]. Overall, ratings were higher on the post-instruction administration than on the pre-instruction administration. There was not a statistically significant main effect for group [F (1, 99) = 2.585, *p* = .111,

$\eta_p^2 = .025$]. Thus, the varying instructional materials (i.e., conventional problems versus worked examples) did not have a significant impact on participants' chi-square self-efficacy ratings.

Perceived knowledge. Perceived knowledge of introductory statistics and perceived knowledge of the chi-square test of independence were both measured using single items, each on ordinal scales. An examination of histograms showed that responses to the item concerning perceived knowledge of introductory statistics was approximately normally distributed at both the pre- and post-instruction points for both groups. Responses to the item concerning perceived knowledge of the chi-square test of independence were highly positively skewed at the pre-instruction administration for both groups. At the post-instruction administration responses were normally distributed for both groups. Due to the ordinal scales of the responses and lack of normality, parametric methods were not employed. Instead, Mann-Whitney U analyses were performed on changes in perceived knowledge ratings.

Changes in perceived knowledge ratings were computed using a simple score change calculation. Pre-instruction ratings were subtracted from post-intervention ratings for each of the two perceived knowledge items. The changes for each type of perceived knowledge were compared between the worked example and conventional problem solving groups. The median changes were the same for the two conditions on both of the perceived knowledge items. Thus, there was not a main effect for group for either item.

For both groups, there was no change in the median rating given to the item concerning perceived knowledge of introductory statistics. A Wilcoxon signed ranks test was performed for all participants combined. Results were not statically significant for the perceived knowledge of introductory statistics item [positive ranks (n = 38; mean = 29.42), negative ranks (mean = 34.79, n = 24), ties = 32; Z = -1.010, *p* = .312]. There was no main effect for administration for the perceived knowledge of introductory statistics item. The median did not change from the pre- to the post-instruction administration.

For both groups, there was a three point increase in the median rating on the item concerning the chi-square test of independence. This was a significant change [positive ranks (n = 79; mean = 49.21), negative ranks (mean = 24.88, n = 12), ties = 9; Z = -7.127, $p < .001$]. There was a main effect for administration for the perceived knowledge of the chi-square test of independence. Scores were higher at the post-instruction administration.

A Mann-Whitney U was used to compare changes in perceived knowledge for the conventional problem solving and worked example groups. These results were not statistically significant. When considering changes perceived knowledge of introductory statistics, the conventional problem solving group had a mean rank of 49.01 (n=46) and the worked examples group had a mean rank of 46.05 (n=48). There was no significant difference between these ranks (U = 1034.500, Z = 0.540, $p = .589$). For the changes in perceived knowledge of the chi-square test of independence variable, the conventional problem solving group had a mean rank of 51.96 (n=49) and the worked examples group had a mean rank of 49.10 (n=51). Again, these results were not statistically significant (U = 1178.000, Z = 0.497, $p = .620$). There were no significant differences between the conventional problem solving and worked example groups in terms of the changes in the ratings to either of the perceived knowledge items.

Demonstrated topic knowledge. The analyses performed on the demonstrated knowledge variable mirrored those performed on the chi-square self-efficacy variable. Descriptive statistics of the composite scores are presented in Table 21. These statistics are reported given scores that have a minimum possible value of 0 and a maximum possible value of 16. The means and standard deviations for each item by group are presented in Table 22. These statistics are reported given scores that have a minimum possible value of 0 and a maximum possible value of 2.

Table 21: Demonstrated Knowledge Composite Scores by Group

Group	n	Pre M	Pre SD	Post M	Post SD
Conventional	49	3.878	2.505	10.449	4.052
Worked Example	51	4.588	2.692	11.137	3.693
Total	100	4.240	2.614	10.800	3.869

Table 22: Demonstrated Knowledge Item Statistics by Condition and Administration

	Conventional Problem Solving				Worked Examples			
	Pre		Post		Pre		Post	
Item Type	M	SD	M	SD	M	SD	M	SD
Identify the assumptions of a χ^2 test of independence	.000	.000	.939	.922	.038	.194	.765	.907
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	.080	.396	.776	.963	.115	.427	.784	.986
Look up the appropriate critical value on a χ^2 table	.680	.957	1.469	.892	.731	.972	1.569	.831
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	.320	.741	1.184	.993	.500	.874	1.451	.901
Calculate row and column totals	.100	.303	1.204	.866	.096	.358	1.275	.750
Calculate expected cell values	1.960	.198	2.000	.000	1.962	.277	1.941	.311
Compute the test statistic for a 2x2 χ^2 test of independence	.380	.780	1.449	.843	.538	.851	1.608	.750
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	.360	.749	1.429	.866	.577	.848	1.745	.560

A repeated measures analysis of variance was performed. Group served as the between groups independent variable. Administration (i.e., pre- and post-instruction) served as a within groups independent variable. The dependent variable was composite demonstrated knowledge score.

The assumption of equal covariance matrices was met [Box's M = 0.870, F (3, 1823999.781) = 0.284, p = .837] as was the assumption of homogeneity of variances which was assessed via Levene's test of equality of error variances [pre-test F (1, 98) = 1.111, *p* = .294; post-test F (1, 98) = 0.270, *p* = .605]. The repeated measures analysis of variance resulted in no statistically significant administration by group interaction [F (1, 98) = 0.001, *p* = .977, $\eta_p^2 < .001$]. There was an overall statistically significant main effect of administration [F (1, 98) = 284.737, *p* < .001, $\eta_p^2 = .744$]. Participants scored higher at the post-instruction administration than at the pre-instruction administration. There was no significant main effect of group [F (1, 98) = 1.720, *p* = .193, $\eta_p^2 = .017$]. These effects are depicted in Figure 6.

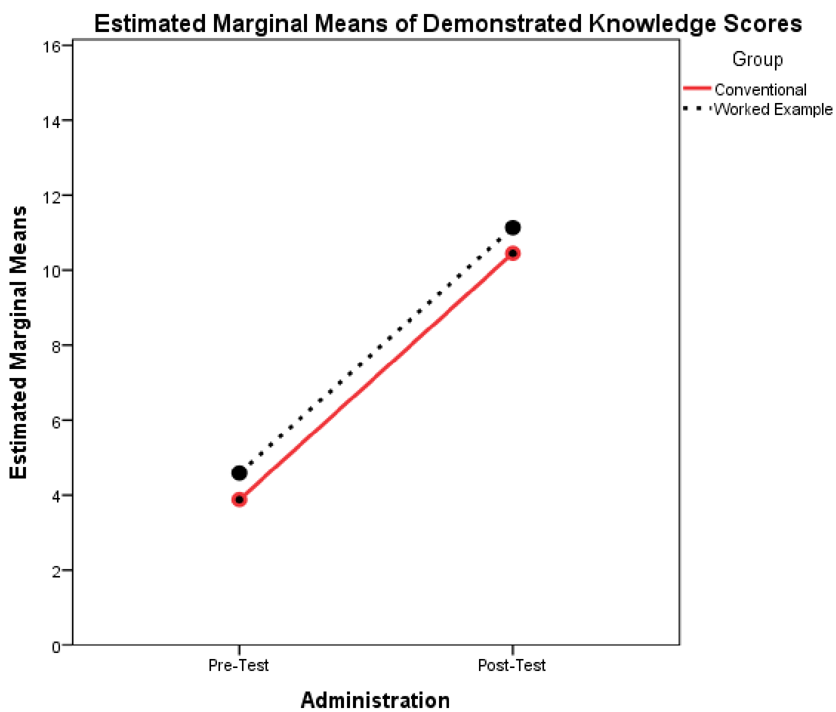


Figure 6: Plot of Margin Means of Demonstrated Knowledge by Group

Research Question 2

To what extent do instructional materials designed to elicit low and high levels of cognitive load impact the relations between self-efficacy, perceived knowledge, and demonstrated knowledge?

First, the Pearson correlation coefficients between self-efficacy and demonstrated knowledge were examined. Both of these variables were measured on interval levels scales. Second, residual gain scores were used to take into account the changes in the two variables from the pre- to post-instruction administrations. Doing so eliminated the correlation between pre-instruction scores and the resulting unstandardized residual gain score (Dimitrov & Rumrill, 2003; Glass & Hopkins, 1996). The use of residual gain scores allowed for comparisons to be made between the conventional problem solving and worked example groups. Finally, relations with the perceived knowledge items were examined using nonparametric techniques, namely Kendall's tau.

Self-efficacy and demonstrated knowledge. The correlation coefficients between pre- and post-instruction chi-square self-efficacy and demonstrated knowledge variables for the conventional problem solving and worked examples group are presented in Table 23. These are the correlation coefficients between the raw data for each variable before adjustments were made to partial out pre-instruction scores. Perceived knowledge was not included in this analysis because it was measured on an ordinal scale.

Table 23: Correlations for Unadjusted Self-Efficacy and Demonstrated Knowledge by Condition

		1	2	3	4
1. Pre-Instruction Chi-Square Self-Efficacy	*r*	---	.153	.109	-.198
	p		.289	.455	.172
	n		50	49	49
2. Pre- Instruction Demonstrated Knowledge	*r*	-.256	---	.202	.305
	p	.067		.164	.033
	n	52		49	49
3. Post- Instruction Chi-Square Self-Efficacy	*r*	.047	.078	---	.498
	p	.738	.585		.000
	n	52	52		49
4. Post- Instruction Demonstrated Knowledge	*r*	-.064	.358	.704	---
	p	.656	.010	.000	
	n	51	51	51	

Values above the diagonal are for the conventional problem solving group. Values below the diagonal are for the worked examples group.

For the conventional problem solving group, the correlation coefficient between chi-square self-efficacy and demonstrated knowledge increased from $r = .153$ at the pre-test to $r = .498$ at the post-test. Using a two-tailed test of statistical significance for dependent correlation coefficients with no variable in common (e.g., Lee & Preacher, 2013), the difference between these two correlation coefficients was approaching statistical significance (z = 1.899, *p* = .058). For the worked example group, the correlation coefficient increased from *r* = -.256 to *r* = .704. This difference was statistically significant when a two-tailed test was performed (z = 5.687, *p* < .001).

The impact of the instructional intervention on chi-square self-efficacy and demonstrated topic knowledge scores was evaluated using residual gain scores. Residual gains were computed for each variable using pre-instruction scores as the predictor variable and post-instruction as the dependent variable; this was performed for the conventional problem solving and worked examples groups separately. Residual gains were used to partial out the effect of the pre-instruction scores on the post-instruction scores (Dimitrov & Rumrill, 2003; Glass & Hopkins, 1996).

Descriptive statistics for the computed unstandardized residuals for the conventional problem solving and worked example groups are given in Table 24. Histograms were also assessed for each variable in each group. No distribution was deemed severely deviant from the normal distribution. There were no outliers that were more than three standard deviations from the mean.

Table 24: Descriptive Statistics for Unstandardized Residuals by Condition

						Skewness		Kurtosis	
	n	Min	Max	M	SD	Stat	SE	Stat	SE
Conventional Problem Solving									
Chi-Square Self-Efficacy	49	-5.211	2.686	0	2.002	-.912	.340	.118	.668
Demonstrated Knowledge	49	-7.522	6.478	0	3.859	-.307	.340	-.715	.668
Worked Example									
Chi-Square Self-Efficacy	52	-3.812	2.194	0	1.719	-.648	.330	-.482	.650
Demonstrated Knowledge	51	-5.885	5.152	0	3.449	-.212	.333	-1.053	.656

The residual gains for chi-square self-efficacy scores were correlated with the residual gains for demonstrated knowledge scores for the two groups separately. For the conventional problem solving condition, the correlation coefficient was r = .489 (p < .001, n = 49, r^2 = .239). For the worked example group, the correlation coefficient was r = .717 (p < .001, n = 51, r^2 = .514). Thus, both groups experienced statistically significant relations between post-instruction chi-square self-efficacy and demonstrated knowledge when partialling out pre-instruction scores.

To compare the correlation coefficients between the two groups, Fisher's Z transformation was performed. The Fisher's Z transformation computes the inverse hyperbolic tangent of the correlation coefficient (i.e., $Z = \tanh^{-1} r$; $\mathrm{Z}_{\mathrm{Conventional}} = 0.535$, $\mathrm{Z}_{\mathrm{Worked\ Example}} = 0.901$). These transformations were then used to compute a z test statistic: $z = (Z_1 - Z_2)/\sigma_{Z_1 - Z_2}$, where $\sigma_{Z_1 - Z_2} = \sqrt{[1/(n_1 - 3)] + [1/(n_2 - 3)]}$ (Glass & Hopkins, 1996, pp. 355-360). A two-tailed test failed to yield statistically significant results (z = 1.777, p = .076). Specifically, there was no significant difference in the correlation coefficients between chi-square self-efficacy and demonstrated topic knowledge residual gain scores for the conventional problem solving and worked example groups.

Perceived knowledge relations. To evaluate correlation coefficients with the perceived knowledge items, Kendall's tau was employed. All correlation coefficients are presented in Table 25. The relations of primary interest were those between perceived knowledge and demonstrated knowledge. For those correlations, each tau was converted to a Pearson's r correlation using the formula $r = sin(.5\pi\tau)$ as defined by Walker (2003; see Figure 7). Then, the Pearson's r correlation coefficients were compared for the conventional problem solving and worked examples groups (see Table 26). The p-values for the difference between the two groups are shown in the last column (p_d); the correlation coefficients for the two groups did not statistically significantly differ on any of the four relations of interest, however three of the correlation coefficients were approaching statistical significance.

Table 25: Kendall's Tau Correlations by Condition

		1	2	3	4	5	6	7	8
1. Pre-Test Chi-Square Self-Efficacy	*Tau*		.137	.286**	.427**	.009	-.203*	.107	.014
	p		.204	.009	<.001	.931	.05	.327	.896
	n		50	48	50	49	49	47	49
2. Pre-Test Demonstrated Knowledge	*Tau*	-.151		.185	.272*	.247*	0.181	.216	.296**
	p	.145		.114	.022	.023	.104	.063	.009
	n	52		48	50	49	49	47	49
3. Pre-Test Perceived Knowledge: Introductory Statistics	*Tau*	-.061	.095		.292*	.228*	0.034	.432**	.224
	p	.571	.400		.015	.038	.761	<.001	.050
	n	49	49		48	47	47	46	47
4. Pre-Test Perceived Knowledge: χ^2 Tests of Independence	*Tau*	.510**	.045	.062		-.003	-.139	.213	.082
	p	<.001	.684	.592		.978	.222	.074	.479
	n	52	52	49		49	49	47	49
5. Post-Test Chi-Square Self-Efficacy	*Tau*	.029	.021	.084	-.051		.312**	.402**	.739**
	p	.770	.840	.437	.626		.002	<.001	<.001
	n	52	52	49	52		49	47	49
6. Post-Test Demonstrated Knowledge	*Tau*	-.050	.274*	.074	-.060	.521**		.006	.273*
	p	.623	.011	.504	.579	<.001		.955	.011
	n	51	51	48	51	51		47	49
7. Post-Test Perceived Knowledge: Introductory Statistics	*Tau*	.090	.104	.057	.004	.384**	.386**		.578**
	p	.403	.363	.626	.972	<.001	<.001		<.001
	n	51	51	48	51	51	51		47
8. Post-Test Perceived Knowledge: χ^2 Tests of Independence	*Tau*	.033	.107	.022	-.025	.646**	.476**	.404**	
	p	.748	.327	.847	.819	<.001	<.001	<.001	
	n	51	51	48	51	51	51	51	

Values above the diagonal are for the conventional problem solving group. Values below the diagonal are for the worked examples condition.

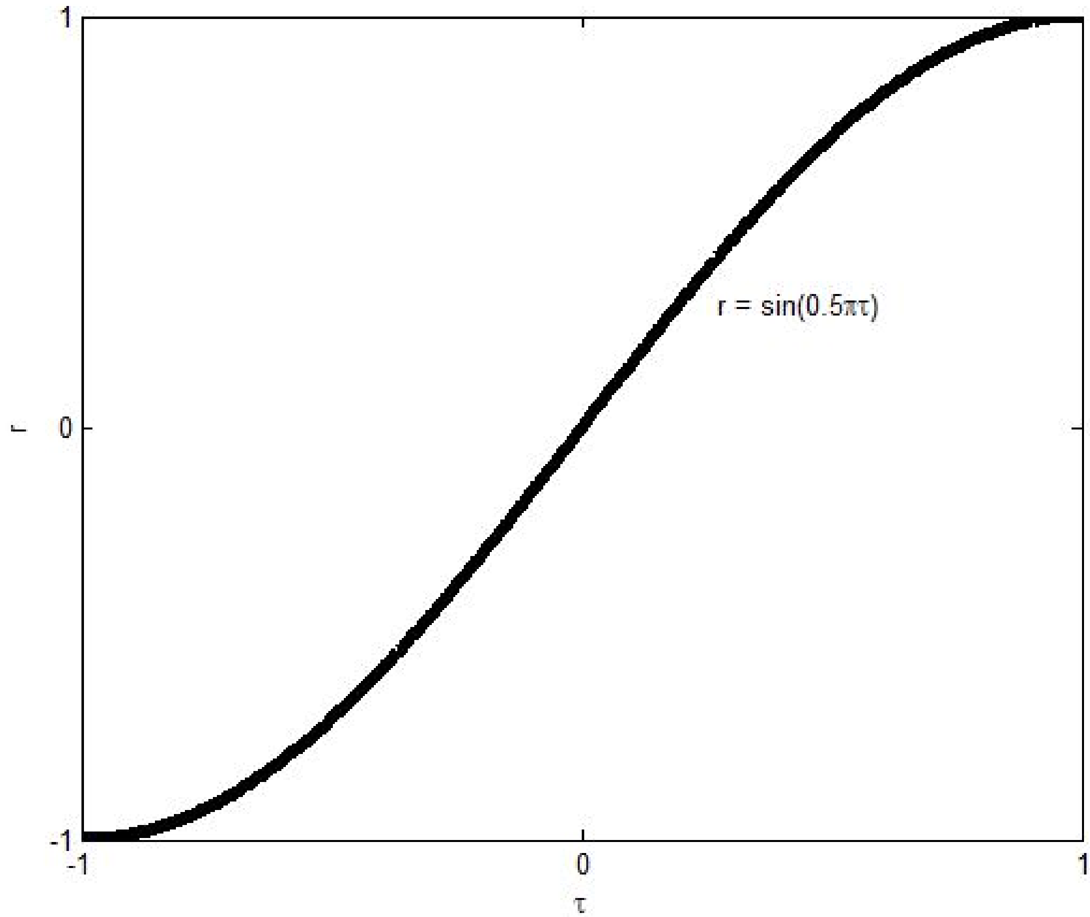


Figure 7: Conversion of Kendall's Tau to Pearson's *r*

Table 26: Pearson's *r* for Perceived and Demonstrated Knowledge by Condition

	Conventional Problem Solving				Worked Example				Difference between Conditions	
	tau	r	p_r	n	tau	r	p_r	n	Z	p
Pre-Instruction										
Introductory Statistics	.185	.287	.048	48	.095	.149	.307	49	0.69	.490
χ^2 Tests of Independence	.272	.414	.003	50	.045	.071	.617	52	1.81	.070
Post-Instruction										
Introductory Statistics	.006	.009	.952	47	.386	.570	<.001	51	1.91	.056
χ^2 Tests of Independence	.273	.416	.003	49	.476	.680	<.001	51	1.87	.062

All correlation coefficients are for the respective perceived knowledge item and the demonstrated knowledge scale score.

Changes in correlation coefficients between the perceived knowledge items and demonstrated knowledge were also evaluated for the conventional problem solving and worked example groups separately. The correlations were considered dependent with no variables in common (e.g., Lee & Preacher, 2013). All correlations were converted from tau to Pearson's *r* for these analyses (e.g., Walker, 2003). For the conventional problem solving group, there was no statistically significant change in the relation between perceived knowledge of introductory statistics and demonstrated knowledge from the pre- to post-test ($\mathrm{r_{pre}} = .287 \rightarrow \mathrm{r_{post}} = .009$, z = -1.491, *p* = .136, $\Delta R^2 = -.082$) nor was there a statistically significant change in the relation between perceived knowledge of the chi-square test of independence and demonstrated knowledge ($\mathrm{r_{pre}} = .414 \rightarrow \mathrm{r_{post}} = .416$, z = 0.011, *p* = .991, $\Delta R^2 = .002$).

In the worked example group, a statistically significant change in the relation between perceived knowledge of introductory statistics and demonstrated knowledge was observed ($\mathrm{r_{pre}} = .149 \rightarrow \mathrm{r_{post}} = .570$, z = 2.392, *p* = .017, $\Delta R^2 = .303$). Similarly, there was a statistically significant change in the relation between perceived knowledge of the chi-square test of independence and demonstrated knowledge in the worked example group ($\mathrm{r_{pre}} = .071 \rightarrow \mathrm{r_{post}} = .680$, z = 3.592, *p* < .001, $\Delta R^2 = .457$).

In summary, correlations with the perceived knowledge items were converted from Kendall's tau to Pearson product moment correlations to allow for statistical comparisons. No statistically significant changes were observed in the conventional problem solving group from the pre- to post-instruction administrations. Statistically significant changes were observed in the worked example condition for the correlations between demonstrated knowledge and both perceived knowledge items.

Summary of Results

The efficacy of the two problem sets (i.e., worked example versus conventional problem solving) was evaluated by comparing participants' ratings of perceived cognitive load. Perceived cognitive load ratings given following the instructional video served as a covariate. The results of the analysis of covariance provided evidence for the efficacy of the practice problems. Perceived cognitive load ratings while problem solving were higher for the conventional problem solving group when compared to the worked examples group. A medium effect size was observed (d = 0.400).

The first research question examined differences between the conventional problem solving and worked example groups in terms of changes in chi-square self-efficacy, perceived knowledge of introductory statistics, perceived knowledge of the chi-square test of independence, and demonstrated topic knowledge of the chi-square test of independence. For the chi-square self-efficacy variable, there was not a main effect for group, but there was a main effect for administration: higher ratings were given on the post-instruction administration when compared to the pre-instruction administration regardless of group. Results were similar for the demonstrated knowledge variable in that there was no main effect for group but there was a main effect for administration with participants scoring higher on the post-instruction administration. Analyses of the perceived knowledge items also found no statistically significant difference between the conventional problem solving and worked examples groups based on the non-parametric analyses which were used to make these comparisons.

The second research question examined changes in the relations between demonstrated topic knowledge and the perceived rating scale variables (i.e., chi-square self-efficacy, perceived knowledge of introductory statistics, perceived knowledge of the chi-square test of independence). To compare the relations between demonstrated knowledge and chi-square self-efficacy at the two administrations and for the two groups, residual gain scores were computed. Pre-instruction scores were used to predict post-instruction scores. These residuals were then used to compute the correlation coefficient between demonstrated knowledge and self-efficacy for each group separately. These correlation coefficients were compared for the two groups. There was no statistically significant difference in the residual gain scores for the conventional problem solving and worked examples conditions.

Because the perceived knowledge items were measured on an ordinal scale, Kendall's tau was first used to evaluate the relations between each perceived knowledge item and demonstrated knowledge. Each tau correlation was converted to a Pearson's *r* correlation which allowed for comparisons. The conventional problem solving group and worked example group were not statistically significantly different from one another for any of the individual correlations. However, there were statistically significant results when examining the changes in the correlations between both perceived knowledge items and demonstrated knowledge from pre- to post-intervention for the worked example group only. For the worked example group, an increase in the correlations between demonstrated knowledge and each of the perceived knowledge items was observed. For the conventional problem solving group no significant change in correlation was observed.

Chapter 5: Discussion

In this final chapter, this research project is summarized and discussed in a number of ways. First, a general overview of the studies is presented. The results of the full study are discussed in terms of the efficacy of the instructional intervention that was employed and in terms of the two primary research questions. Limitations of the present study are summarized and practical implications for post-secondary statistics education are discussed. Finally, suggestions for future research are given.

General Overview

The purpose of this research was to examine learners' self-efficacy, perceived knowledge, and demonstrable topic knowledge in conjunction with instructional interventions designed to elicit lower or higher levels of perceived cognitive load. This was based on the idea that students' behaviors and perceptions of their own capabilities are influenced by experiences, in this case the experience of working through practice problems. The relations between behavioral and perceptional variables of interest were operationalized using measures of demonstrated knowledge and self-efficacy within the introductory statistics topic of the chi-square test of independence. Perceived knowledge of introductory statistics and perceived knowledge of the chi-square test of independence were also measured. The experience of being exposed to practice problems employing either conventional problem solving or worked example methods served as the between groups variable in the primary research study. The practice problem sets sought to elicit lower and higher levels of perceived cognitive load, respectively.

The chi-square test of independence was selected as the topic of focus because it is an introductory topic that is not typically taught until later in the course from which participants were recruited. The chi-square test of independence requires only a minimal level of background knowledge of statistics and involves relatively basic mathematical operations. It was also selected because its procedures can be taught in a relatively short period of time.

While there was existing literature concerning self-efficacy and existing literature concerning cognitive load, as reviewed in Chapter 2, there was little previous work that had examined the two constructs together, particularly in the area of statistics education. Thus, the present research sought to address underdeveloped areas within the field relative to these constructs. There were two primary research questions that were addressed by the full research study:

1. To what extent do instructional materials designed to elicit low and high levels of cognitive load impact self-efficacy, perceived knowledge, and demonstrated knowledge?
2. To what extent do instructional materials designed to elicit low and high levels of cognitive load impact the relations between self-efficacy, perceived knowledge, and demonstrated knowledge?

Discussion of Results

Efficacy of intervention. The effect of the worked example group's practice set versus the conventional problem solving group's practice set was evaluated in the full study. An analysis of covariance was used to compare the perceived cognitive load ratings given by participants in the worked example and conventional problem solving groups. Perceived cognitive load ratings given following the instructional video served as a covariate. Those baseline ratings were a statistically significant covariate in the analysis of covariance. That is, by including the baseline ratings in the model, some variance due to preexisting individual differences could be accounted for.

There was a statistically significant difference between the mean perceived cognitive load ratings given by the two groups. Participants in the conventional problem solving group gave higher perceived cognitive load ratings on average than participants in the worked example group. The effect size for the difference, however, was only small to moderate. While this provides evidence for the efficacy of the instructional intervention and is consistent with the existing research (e.g., Paas, 1992), a larger effect size would be desirable.

The lack of a larger effect size for this analysis may have been due to the minimal differences between two groups' instructional interventions. The only difference between the materials that the two groups were exposed to was the format of the first practice problem set. If the worked example group had been exposed to more than one worked example, then the group's perceived cognitive load may have been lower relative to that of the conventional problem solving group.

The result of this analysis was also influenced by the method used to calculate perceived cognitive load while problem solving. In the second pilot study and in the full study, perceived cognitive load while problem solving was computed by finding the sum of the responses to the mental effort scale after the three practice problem sets. For the worked example group, this calculation method meant that ratings for the initial worked example and the two subsequent conventional problem sets were combined. In doing so, differences by practice set may have been masked.

Research question 1. The first research question addressed changes in self-efficacy, perceived knowledge, and demonstrated knowledge from before to after the instructional intervention. The worked example and conventional groups were compared. There were no interaction effects between group and administration (before and after the instructional intervention). In the following sections, the main effects for group and administration are discussed.

Main effect of group. In relation to the first research question, the results of the full study did not find statistical evidence of differences in self-efficacy, perceived knowledge, or demonstrated topic knowledge for the worked example and conventional problem solving groups. In other words, there were no main effects for group for any of the analyses performed.

The lack of statistical significance for the comparison of demonstrated knowledge scores for the two groups was surprising. This result was not consistent with the findings of the previous studies of

Sweller and Cooper (1985) or Paas (1992) in which the worked example groups outperformed other conditions on knowledge-related measures. A factor contributing to the results of the present study may have been task difficulty. The demonstrated knowledge measure only covered the 2x2 chi-square test of independence. Items of greater difficulty or items measuring far transfer, such as a 3x3 chi-square test of independence, may have altered the findings.

Main effect of administration. While there were no main effects for condition, there were some main effects when comparing the pre- and post-instructional intervention administrations. Specifically, there were main effects for administration for three of the four variables that were analyzed. For the chi-square self-efficacy, perceived knowledge of the chi-square test of independence, and demonstrated topic knowledge variables there were statistically significant increases in scores from the pre- to post-instructional intervention administrations. Effect sizes associated the main effects of administration were computed for the quantitative variables of chi-square self-efficacy and demonstrated topic knowledge. In both cases, the effect sizes were very large.

The very large effect size for the improvement in demonstrated topic knowledge scores provides evidence for the value of this brief instructional intervention. Though the study was relatively short, consisting of a video of approximately 11 minutes and three practice problem sets, statistically significant and practically significant increases in participants' demonstrated topic knowledge scores were observed. Looking at the individual items on the demonstrated topic knowledge form, the tasks for which participants experienced the greatest improvements were those associated with: (a) making decisions to reject or fail to reject the null hypothesis; (b) computing the chi-square test statistic; and (c) computing expected values. The tasks for which participants may need more instruction, as indicated by having the lowest post-test scores, were: (a) writing appropriate hypotheses and (b) identifying assumptions. This information can be used to inform the design of instructional materials in the future. The items on which the largest gains were observed were all related to the content of the practice problems. The post-test items with which participants struggled the most were neither computational nor explicitly overviewed in the practice problems.

Participants, regardless of the type of practice problem sets they received, on average saw an increase in scores for the variables directly related to the content of the instructional materials from the pre- to post-instructional intervention administrations. But, there was not a statistically significant change in responses to the perceived knowledge of introductory statistics item from pre- to post-instructional intervention administrations. Of the four variables compared by administration, perceived knowledge of introductory statistics was the least closely related to the instructional intervention materials and was the broadest. The chi-square test of independence is only one small part of the set of topics in introductory statistics.

Research question 2. The self-efficacy and demonstrated knowledge variables were approximately normal and measured on scales that were treated interval-level. Therefore, Pearson's correlation coefficients were used along with analyses of residual gains to examine the relations between the two variables and changes in the two variables from the pre- to post-instructional intervention administrations for participants assigned to the two conditions. Perceived knowledge was measured with two ordinal-level items. Thus, the analyses of the relations with the perceived knowledge items employed nonparametric techniques, namely Kendall's tau. Tau correlations were converted to Pearson's *r* for some between and within group comparative analyses. Here, the results of the analyses between self-efficacy and demonstrated topic knowledge are reviewed. This is followed by a discussion of the analyses between the perceived knowledge items and demonstrated topic knowledge.

Self-efficacy and demonstrated topic knowledge.

Changes in correlation coefficients. Pearson correlation coefficients between self-efficacy and demonstrated knowledge increased from pre- to post-instruction for both groups. This change suggests an increase in self-efficacy calibration. Calibration is the relationship between perceptions of one's abilities and that ability demonstrated in an observable manner (Glenberg, Sanocki, Epstein, & Morris, 1987; see also Alexander, 2013; Bembenutty, 2009). According to Bembenutty, “the metacognitive processes involved in self-regulation include a comparison between one’s judgment of learning and one’s actual performance. In other words, learners’ calibration between confidence of knowing and actual performance is an essential feature of self-regulation” (p. 562).

There are practical implications to self-efficacy calibration. For example, according to Glenberg, Sanocki, Epstein, and Morris (1987), “In preparing for a test of learning, a rational strategy is to study until one believes that the material is learned. Studying for less time is risky; studying for more time may be wasteful” (p. 119). In the setting of an introductory statistics course, this calibration is important because it may be related to students’ study behaviors. Students who overestimate their abilities may underestimate the studying they must do to reach their goals for performance on an assessment such as an exam. Students who underestimate their abilities may experience anxiety or other affects associated with low self-efficacy. However, it is also known that self-efficacy is positively correlated with persistence and effort (Lent et al., 1984; Schunk, 1981). As such, having high self-efficacy, regardless of ability, may lead to greater persistence and effort. There is still much research to be done on self-efficacy calibration in educational settings.

Given the results of this study, self-efficacy calibration may differ for individuals experiencing varying levels of perceived cognitive load. This is evidenced by the greater change in the relationship between self-efficacy and demonstrated topic knowledge that was observed in the worked example group when compared to the conventional problem solving group. While these results were not statistically significant, more research on the relation between experiences and self-efficacy calibration is needed before conclusions can be made.

Residual gain analyses. To evaluate the differences in the relations between self-efficacy and demonstrated knowledge for participants in the two conditions, residual gain scores were used. Residual gains were computed for each variable using pre-instruction scores as the predictor variable and post-instruction as the dependent variable. These residual gain computations were performed for the conventional problem solving and worked example groups separately. By correlating the residual gains, the effects of the pre-test scores were eliminated within each variable (Dimitrov & Rumrill, 2003; Glass & Hopkins, 1996).

The residual gains for chi-square self-efficacy scores were correlated with the residual gains for demonstrated knowledge scores for the worked example and conventional problem solving groups separately. For both groups, the relation between post-instruction chi-square self-efficacy and demonstrated knowledge when partialling out pre-instruction scores was statistically significant and positive. The correlation coefficient was stronger in the worked example group, however the difference was not statistically significant. The positive correlation coefficients between the residual gains for the self-efficacy and demonstrated knowledge variables for both groups signify a relation between improvements observed on each of the two scales. For example, participants who experienced a greater than expected increase in their demonstrated knowledge tended to also experience a greater than expected increase in their self-efficacy.

Relations with perceived knowledge. The second research question was also assessed through the comparisons of correlation coefficients between demonstrated knowledge and the perceived knowledge variables. Because the perceived knowledge items were measured on an ordinal scale, Kendall's tau was first used as a measure of the bivariate relations. In order to make group comparisons and to examine changes in correlations, each tau correlation coefficient was converted to a Pearson's *r* correlation coefficient. There were no statistically significant differences between the correlation coefficients for the worked example and conventional problem solving groups. When changes in relations were examined, there were no statistically significant changes in the correlation coefficients from pre- to post-instruction administrations for the conventional problem solving group on either perceived knowledge item. For the worked example group, there were statistically significant changes for the correlations between demonstrated knowledge and both perceived knowledge of introductory statistics and perceived knowledge of the chi-square test of independence.

These results suggest possible improvements in calibration for the worked example group. These analyses, however, assess perceived knowledge calibration as opposed to self-efficacy calibration. Due to the ordinal nature of the perceived knowledge items, and the numerous transformations that had to be performed, these analyses should be interpreted with some caution. In the future, an interval-level measure of perceived knowledge should be employed to allow for parametric comparisons without having to transform non-parametric correlation coefficients to Pearson product moment correlation coefficients before analyses can be performed. These results, however, are promising in that they suggest that those who received the worked example experienced greater improvement than those who received the conventional problem solving set.

Limitations

Sample characteristics. The relatively small samples sizes in the pilot studies were limitations. Additionally, the homogeneousness of the samples was a limitation. For example, all participants in the first pilot study were graduate students in education-related fields and all participants in the second pilot study and the full study were undergraduate students enrolled in introductory statistics courses. All participants in the full study were between the ages of 18 and 24 and all were recruited from the same introductory statistics course.

Differences between the samples utilized in the pilot studies, particularly the first pilot study, and the full study were also a limitation. Participants in the first pilot study were graduate students while participants in the other studies were undergraduate students. Thus, the sample that was used to pilot the instruments was drawn from a different population than the sample from the full study.

Effect size of perceived cognitive load difference. The lack of a large effect of the two practice problem sets on participants' ratings of perceived cognitive load was also a limitation. While in the full study the intervention did produce a statistically significant difference in perceived cognitive load ratings following the practice problem sets, a larger effect size for this difference is desirable. An intervention that elicited a greater difference in terms of perceived cognitive load may have changed the results of the analyses for the following research questions which compare the conventional problem solving and worked example groups.

In the present study, perceived cognitive load was manipulated by varying the instructional materials. In turn, only extraneous cognitive load was intentionally altered. As previously suggested, the intrinsic load of the content may not have been great enough to cause participants to experience cognitive load when faced with the problem solving experiences.

Test-retest effect. Alternate forms of the demonstrated knowledge measure were used to prevent test-retest effects. However, the chi-square self-efficacy and perceived knowledge items did not change from the pre- to post-instructional intervention administrations. Additionally, the self-efficacy and perceived knowledge scales were administered before the demonstrated knowledge items. As seen in the first study, the act of completing the demonstrated knowledge items could have impacted participants' self-efficacy ratings. Therefore, the post-instructional intervention self-efficacy ratings may have been influenced not only by the instructional intervention, but also the initial exposure to the demonstrated knowledge items.

Implications for Post-Secondary Statistics Education

Instructional intervention. While the results of the full study do not show a statistically significant difference between the worked example and conventional problem solving groups in terms of demonstrated topic knowledge scores, the results do show that the instructional intervention was related to improved scores. Regardless of whether the worked example or conventional problem solving set of practice questions was administered, an average increase in demonstrated topic knowledge was observed. In the full study, the effect size for the main effect of administration on demonstrated topic knowledge scores was very large. Participants' scored about two standard deviations higher on the post-instructional intervention measure of demonstrated topic knowledge than they did on the pre-instructional intervention measure. Thus, even a short instructional intervention such as the one featured in the full study, comprised of an 11 minute video and three practice problem sets, may be related to increases in demonstrated topic knowledge. Interventions such as the one used in this dissertation may be effective and efficient instructional methods for some introductory statistics topics.

The effectiveness of the instructional intervention can also be examined by individual task. As previously stated, the tasks for which participants experienced the greatest improvements were those associated with: (a) making decisions to reject or fail to reject the null hypothesis; (b) computing the chi-square test statistic; and (c) computing expected values. These are all tasks that were a part of the practice problem activities. The tasks that participants had the lowest post-instruction scores on were those concerning: (a) writing appropriate hypotheses and (b) identifying assumptions. These were tasks that were not explicitly a part of the practice problem activities. Instructional interventions used in the classroom should be sure to emphasize all tasks of importance in the practice activities. Doing so gives students repeated practice with applying these concepts which may lead to greater increases in demonstrated knowledge.

In the present study, the instructional intervention was presented entirely online. This allowed for very high consistency in that all participants were presented with the same instruction. In a classroom setting, this would mean that the same instructional experience would be available to all students. Additionally, participants worked through the materials at their own paces. They were able to re-watch the instructional video if they chose to do so. In the practice problem sets, students who answered the questions correctly were able to continue through the study quickly without being required to view information that they already knew. Students who could not answer the questions correctly were given two extra trials to fix their answers; if they were still not able to answer the question correctly they were given the solution with an explanation.

Self-efficacy calibration. A relation between experience and self-efficacy calibration was evident in both the first pilot study and the full study. In both cases, the correlation between self-efficacy and demonstrated topic knowledge was low at the beginning of the study. This signifies low initial self-efficacy calibration. That is, students were not good at judging their own knowledge at the beginning of the studies.

The correlation coefficient between self-efficacy and demonstrated knowledge increased from the beginning to end of both the first pilot study and the full study. For the first pilot study, the relation between self-efficacy and demonstrated knowledge increased despite the lack of an instructional intervention. Demonstrated knowledge scores in that study, however, were composite scores from both the first and second administrations, so results should be interpreted with caution. In the full study, when analyzing the data provided by all participants combined, the correlation coefficient between self-efficacy and demonstrated knowledge increased from the beginning of the study to the end of the study. In both cases, improvements in self-efficacy calibration were observed.

In the classroom, self-efficacy calibration may be improved by providing students with instruction and opportunities to practice. It is unclear from the results of this study what caused the change in self-efficacy calibration. It may have been related to the instructional intervention, the act of completing the chi-square self-efficacy scale, the act of completing the demonstrated knowledge items, or a combination of these events. That is, given this study, it is not possible to identify the precise cause of the changes in self-efficacy and self-efficacy calibration. Having students think about their abilities by completing a self-efficacy survey may have an impact on their self-efficacy calibration. Given that the present study did not explore the impact of being presented with the survey or the demonstrated knowledge items, instructors may consider presenting students with similar measures. For example, they may choose to incorporate self-efficacy scales and practice tests into their instruction.

Suggestions for Future Research

Given the discussion of results, limitations, and educational implications, a number of suggestions for future research have been made. For example, future research in this area should seek to increase the difference between the low and high cognitive load conditions. Increasing the effect size elicited by the intervention may increase the power of the comparative analyses. This may be done by exploring the use of topics that are more difficult, thus increasing intrinsic cognitive load. The design of the present study included materials to attempt experimental manipulation of extraneous cognitive load only. The manipulation of intrinsic cognitive load may also be of interest as it would mimic the experiences of students being presented with material of varying difficulty levels.

Future research should examine self-efficacy calibration and its impact on students' behaviors. For example, relations between self-efficacy calibration and study habits may be of interest. Because an emphasis on practical applications is desirable, interventions that lead to improved self-efficacy calibration should be examined. As previously stated, according to Glenberg et al. (1987), "In preparing for a test of learning, a rational strategy is to study until one believes that the material is learned. Studying for less time is risky; studying for more time may be wasteful" (p. 119). Interventions that improve self-efficacy calibration may be used to increase students' studying efficiency in that students can make more informed decisions pertaining to their studying behaviors.

In the present studies, the act of completing a self-efficacy survey may impact self-efficacy calibration. As seen in the first pilot study, the correlation coefficient between self-efficacy and demonstrated knowledge increased from the first to second administration of the instruments despite the lack of an instructional intervention. The impact of being exposed to measures of self-efficacy and demonstrated knowledge should be examined in future studies that seek to develop effective educational interventions for improving self-efficacy calibration.

Finally, from a methodological perspective, future studies should use directional hypotheses when possible and appropriate. Due to the minimal previous research on self-efficacy and demonstrated knowledge in conjunction with cognitive load, in the present studies, non-directional hypotheses were proposed. Directional analyses, however, would lead to greater statistical power.

Conclusion

In summary, through the review of literature and empirical results, this dissertation project supports the use of worked examples when teaching introductory statistics topics such as the chi-square test of independence. While many of the comparative analyses in the present study were not statistically significant, some were approaching statistical significance and none were in favor of the conventional problem solving condition. These results, in addition to the ample existing research on worked examples, support the use of worked examples in the introductory statistics classroom.

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1–3.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, *61*(3), 315–343.

Amadieu, F., van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cogntive load, and learning. *Learning and Instruction*, *19*(5), 376–386.

Artino, A. R., & McCoach, D. B. (2008). Development and initial validation of the Online Learning Value and Self-Efficacy Scale. *Journal of Educational Computing Research*, *38*(3), 279–303.

Ayres, P. (2006a). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology*, *20*(3), 287–298. http://doi.org/DOI: 10.1002/acp.1245

Ayres, P. (2006b). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389–400. http://doi.org/10.1016/j.learninstruc.2006.09.001

Ayres, P., & Paas, F. (2012). Cognitive load theory: New directions and challenges. *Applied Cognitive Psychology*, *26*(6), 827–832.

Bandura, A. (1977a). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, *84*(2), 191–215.

Bandura, A. (1977b). *Social learning theory*. Englewoods Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (1978). Reflections on self-efficacy. *Advances in Behavioral Research and Therapy*, *1*, 237–269.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147.

Bandura, A. (1986). *Social foundations of thought and action*. Englewoods Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (1992). Exercise of personal agency through the self-efficacy mechanism. In R. Schwarzer (Ed.), *Self -Efficacy: Thought Control of Action* (pp. 3–38). Washington, DC: Hemisphere Publishing Co.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman and Company.

Bandura, A. (2007). Much ado over a faulty conception of perceived self-efficacy grounded in faulty experimentation. *Journal of Social and Clinical Psychology*, *26*(6), 641–658.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, *38*(1), 9–44.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology*, *45*(5), 1017–1028.

Bandura, A., & Jourden, F. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*, *60*(6), 941–951.

Bandura, A., Reese, L., & Adams, N. E. (1982). Microanalysis of action and fear arousal as a function of differential levels of perceived self-efficacy. *Journal of Personality and Social Psychology*, *43*(1), 5–21.

Bartsch, R. A., Case, K. A., & Meerman, H. (2012). Increasing academic self-efficacy in statistics with live vicarious experience presentation. *Teaching of Psychology*, *39*(2), 133–136.

Beckmann, J. F. (2010). Taming a beast of burden- On some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction*, *20*, 250–264.

Bembenutty, H. (2009). Three essential components of college teaching: Achievement calibration, self-efficacy, and self-regulation. *College Student Journal*, *43*(2B), 562–570.

Bruning, R. H., Schraw, G. J., Norby, M. M., & Ronning, R. R. (2004). *Cognitive psychology and instruction* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Brünken, R., Plass, J. L., & Moreno, R. (2010). Current issues and open questions in cognitive load research. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 253–272). New York: Cambridge University Press.

Cashin, S. E. (1999). *Effects of mathematics self-concept, perceived self-efficacy, and attitudes toward statistics on statistics achievement*. Southern Illinois University at Carbondale.

Cashin, S. E. (2001). *Effects of mathematics self-concept, perceived self-efficacy, and attitudes toward statistics on statistics achievement*. Southern Illinois University.

Chiesi, F., Primi, C., & Carmona, J. (2011). Measuring statistics anxiety: Cross-country valdiity of the Statistical Anxiety Scale (SAS). *Journal of Psychoeducational Assessment*, *29*(6), 559–569.

Coladarci, T., Cobb, C. D., Minium, E. W., & Clarke, R. B. (2011). *Fundamentals of statistical reasoning in education* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Ltd.

Connor, J. P., George, S. M., Gullo, M. J., Kelly, A. B., & Young, R. M. (2011). A prospective study of alcohol expectancies and self-efficacy as predictors of young adolescent alcohol misuse. *Alcohol and Alcoholism*, *46*(2), 161–169.

Davidson, J. (2003). *Bandura's social cognitive theory: An introduction*. Davidson Films, Inc.

DeTure, M. (2004). Cognitive style and self-efficacy: Predicting student success in online distance education. *American Journal of Distance Education*, *18*(1), 21–38.

Dimitrov, D. M., & Rumrill, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*, 159–165. Retrieved from http://www.researchgate.net/publication/10826237_Pretest-posttest_designs_and_measurement_of_change/links/0046352385b995ed25000000

Dodge, Y. (2008). Chi-square test of independence. In *The Concise Encyclopedia of Statistics* (pp. 79–82).

Dodson, C. H., & Lewallen, L. P. (2011). Nursing students' perceived knowledge and attitudes toward genetics. *Nurse Education Today*, *31*(4), 333–339.

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, *28*(2), 161–186.

Garner, R., & Gillingham, M. G. (1991). Topic knowledge, cognitive interest, and text recall: A microanalysis. *The Journal of Experimental Education*, *59*(4), 310–319.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston, MA: Allyn and Bacon.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*(2), 119–136.

Gordon, S. (2004). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal*, *3*(1), 40–59. Retrieved from http://www.stat.auckland.ac.nz/~iase/serj/SERJ3(1)_gordon.pdf

Hall, S., & Vance, E. A. (2010). Improving self-efficacy in statistics: Role of self-explanation & feedback. *Journal of Statistics Education*, *18*(3). Retrieved from www.amstat.org/publications/jse/v18n3/hall.pdf

Hsiao, C.-Y., Van Riper, M., Lee, S.-H., Chen, S.-J., & Lin, S.-C. (2011). Taiwanese nursing students' perceived knowledge and clinical comfort with genetics. *Journal of Nursing Scholarship*, *43*(2), 125–132.

Jung, M. E., & Brawley, L. R. (2011). Exercise persistence in the face of varying exercise challenges: A test of self-efficacy theory in working mothers. *Journal of Health Psychology*, *16*(5), 728–738.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539.

Kalyuga, S. (2012). Interactive distance education: A cognitive load perspective. *Journal of Computing in Higher Education*, *24*, 182–2018. http://doi.org/10.1007/s12528-012-9060-4

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*(1), 32–31.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*(1), 126–136. http://doi.org/10.1037/0022-0663.92.1.126

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, *38*(3), 209–215.

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimor skills. *Educational Psychology Review*, *24*(2), 313–337.

Lahl, M., Modic, M. B., & Siedlecki, S. (2013). Percieved knowledge and self-confidence of pediatric nurses as patient educators. *Clinical Nurse Specialist*, *27*(4), 188–193.

Lane, A. M., Hall, R., & Lane, J. (2004). Self-efficacy and statistics performance among sport studies students. *Teaching in Higher Education*, *9*(4), 435–461.

Leahy, W., & Sweller, J. (2008). The imagination effect increases with an increased intrinsic cognitive load. *Applied Cognitive Psychology*, *22*, 273–283. http://doi.org/10.1002/acp.1373

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with no variable in common. Retrieved from http://quantpsy.org/corrtest/corrtest3.htm

Lent, R. W., Brown, S. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psychology*, *31*(3), 356–362.

Leonard, D. C. (2002). *Learning theories, A to Z*. Westport, CT: Greenwood Press.

Leppink, J., Broers, N. J., Imbos, T., van der Vleuten, C. P. M., & Berger, M. P. F. (2012). Self-explanation in the domain of statistics: An expertise reversal effect. *Higher Education*, *63*(6), 771–785.

Lim, C. K. (2001). Computer self-efficacy, academic self-concept, and other predictors of satisfaction and future participation of adult distance learners. *American Journal of Distance Education*, *15*(2), 41–51.

Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology*, *69*(2), 241–251.

Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *The American Statistician*, *54*(3), 196–206.

Low, R., Jin, P., & Sweller, J. (2011). How research in educational psychology has contributed to instructional procedures. In C. M. Rubie-Davies (Ed.), *Educational psychology: Concepts, research, and challenges* (pp. 87–101). New York: Routledge.

Maradiegue, A., Edwards, Q. T., Seibert, D., Macri, C., & Sitzer, L. (2005). Knowledge, percetpions, and attitudes of advanced practice nursing students regarding medical genetics. *Journal of the American Academy of Nurse Pracititioners*, *17*(11), 472–479.

Mayer, R. E. (2013). Incorporating motivation into multimedia learning. *Learning and Instruction*, *29*, 171–173.

McQuaid, J. W. (2010). Using cognitive load to evaluate participation and design of an asynchronous course. *American Journal of Distance Education*, *24*(4), 177–194.

Miltiadou, M., & Yu, C. H. (2000). *Validation of the Online Technologies Self-Efficacy Scale (OTSES)*. Tempe, AZ. Retrieved from http://www.creative-wisdom.com/pub/efficacy.pdf

Moreno, R., Reisslein, M., & Ozogul, G. (2009). Optimizing worked-example instruction in electrical engineering: The role of fading and feedback during problem-solving practice. *Journal of Engineering Education*, *98*(1), 83–92.

Murphy, P. K., & Alexander, P. A. (2004). Persuasion as a dynamic, multidimensional process: An investigation of individual and intraindividual differences. *American Education Research Journal*, *41*(2), 337–363.

Murphy, P. K., & Mason, L. (2006). Changing knowledge and changing beliefs. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (pp. 305–326). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments- a comprehensive review of the literature. *Teaching in Higher Education*, *8*(2), 195–209.

Ozer, E. M., & Bandura, A. (1990). Mechanisms governing empowerment effects: A self-efficacy analysis. *Journal of Personality and Social Psychology*, *58*(3), 472–486.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent development. *Educational Psychologist*, *38*(1), 1–4. http://doi.org/10.1207/S15326985EP3801_1

Paas, F., Tuovinen, J. E., van Merriënboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, *53*(3), 25–34.

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, *16*(2), 87–91.

Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive Load Theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*, 115–121.

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122–133.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578.

Pastore, R. (2012). The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education*, *58*(1), 641–651.

Penn State University Libraries. (2014). LionSearch. Retrieved from http://www.libraries.psu.edu/psul/lionsearch.html#general-information

Pinter, B., Matchock, R. L., Charles, E. P., & Balch, W. R. (2014). A cross-sectional evaluation of student achievement using standardized and performance-based tests. *Teaching of Psychology*, *41*(1), 20–27. http://doi.org/10.1177/0098628313514174

Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010a). *Cognitive Load Theory*. New York: Cambridge University Press.

Plass, J. L., Moreno, R., & Brünken, R. (2010b). Introduction. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 1–6). New York: Cambridge University Press.

Pociask, F. D., & Morrison, G. (2004). The effects of split-attention and redundancy on cognitive load when learning cognitive and psychomotor tasks. In *Association for Educational Communications and Technology*. Chicago, IL. Retrieved from http://files.eric.ed.gov/fulltext/ED485075.pdf

Puzziferro, M. (2008). Online technologies self-efficacy and self-regulated learning as predictors of final grade and satisfaction in college-level online courses. *American Journal of Distance Education*, *22*(2), 72–89.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*(1), 1–29.

Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 91–108). New York: Cambridge University Press.

Rey, G. D., & Fischer, A. (2013). The expertise reversal effect concerning instruction explanations. *Instructional Science*, *41*(2), 407–429.

Rutledge, S. E., Siebert, D. C., Chonody, J., & Killian, M. (2011). Information about human sexuality: Sources, satisfaction, and perceived knowledge among college students. *Sex Education*, *11*(4), 471–487.

Schoor, C., Bannert, M., & Brünken, R. (2012). Role of dual task design when measuring cognitive load during multimedia learning. *Educational Technology, Research and Development*, *60*(5), 753–768.

Schunk, D. H. (1981). Modeling and attributional effects on children’s achievement: A self-efficacy analysis. *Journal of Educational Psychology*, *73*(1), 93–105.

Schunk, D. H. (2012). Social cognitive theory. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, J. Sweller, & J. Brophy (Eds.), *APA Educational Psychology Handbook* (pp. 101–124). Washington, DC: American Psychological Association.

Schunk, D. H., & Usher, E. L. (2011). Assessing self-efficacy for self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 282–297). New York: Routledge.

Senbanjo, R., Wolff, K., Marshall, E. J., & Strang, J. (2009). Persistence of heroin use despite methadone treatment: Poor coping self-efficacy predicts continued heroin use. *Drug and Alcohol Review*, *28*(6), 608–615.

Shen, D., Cho, M.-H., Tsai, C.-L., & Marra, R. (2013). Unpacking online learning experiences: Online learning self-efficacy and learning satisfaction. *The Internet and Higher Education*, *19*, 10–17.

Simmering, M. J., Posey, C., & Piccoli, G. (2009). Computer self-efficacy and motivation to learn in a self-directed online course. *Decision Sciences Journal of Innovative Education*, *7*(1), 99–121.

Spence, D. J., & Usher, E. L. (2007). Engagement with mathematics courseware in traditional and online remedial learning environments: Relationship to self-efficacy and achievement. *Journal of Educational Computing Research*, *37*(3), 267–288.

Stadler, M. A. (1989). On learning complex procedural knowledge. *Journal Of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1061–1069.

Sweller, J. (1976). The effect of task complexity and sequence on rule learning and problem solving. *British Journal of Educational Psychology*, *67*(4), 553–558.

Sweller, J. (2006). How the human cognitive system deals with complexity. In J. Elen & R. E. Clark (Eds.), *Handling Complexity in Learning Environments: Theory and Research* (pp. 13–25). Amsterdam, Netherlands: Elsevier, Inc.

Sweller, J. (2008). Cognitive load theory and the use of educational technology. *Educational Technology*, *48*(1), 32–35.

Sweller, J. (2010a). Cognitive Load Theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 29–47). New York: Cambridge University Press.

Sweller, J. (2010b). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123–138.

Sweller, J., Ayres, P., & Kalyuga, S. (2011a). Measuring cognitive load. In J. Sweller, P. Ayres, & S. Kalyuga (Eds.), *Cognitive Load Theory* (pp. 71–85). Springer.

Sweller, J., Ayres, P., & Kalyuga, S. (2011b). The redundancy effect. In J. Sweller, P. Ayres, & S. Kalyuga (Eds.), *Cognitive Load Theory* (pp. 141–154). Springer.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review1*, *10*(3), 251–296.

Tormala, Z. L., & Petty, R. E. (2007). Contextual contrast and perceived knowledge: Exploring the implications for persuasion. *Journal of Experimental Social Psychology*, *43*(1), 17–30.

Uysal, M. P. (2013). Towards the use of a novel method: The first experiences on measuring the cogntiive load of learned programming skills. *Turkish Online Journal of Distance Education*, *14*(1), 166–184.

Van Gog, T., Kirschmer, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favor of repeated measures. *Applied Cognitive Psychology*, *26*(6), 833–839.

Vasile, C., Marhan, A.-M., Singer, F. M., & Stoicescu, D. (2011). Academic self-efficacy and cognitive load in students. *Procedia, Social and Behavioral Sciences*, *12*, 478–482.

Walker, D. A. (2003). JMASM9: Converting Kendall's tau for correlational or meta-analytic analyses. *Journal of Modern Applied Statistical Methods*, *2*(2), 525–530. Retrieved from http://www.cedu.niu.edu/~walker/personal/Walker Kendall's Tau.pdf

Wang, C.-H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy and course outcomes in online learning. *Distance Education*, *34*(3), 302–323.

Wang, Q., Yang, S., Liu, M., Cao, Z., & Ma, Q. (2014). An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, *62*, 1–10.

Weinberg, R. S., Gould, D. R., & Jackson, A. (1979). Expections and performance: An empirical test of Bandura's self-efficacy theory. *Journal of Sport Psychology*, *1*, 320–331.

Woolfolk Hoy, A., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, *21*(4), 343–356.

Zare, H., Rastegar, A., & Hosseini, S. M. D. (2011). The relation among achievement goals and academic achievement in statistics: The mediating role of statistics anxiety and statistics self-efficacy. *Procedia, Social and Behavioral Sciences*, *30*, 1166–1172.

Zheng, R., & Cook, A. (2012). Solving complex problems: A convergent approach to cognitive load measurement. *British Journal of Educational Technology*, *43*(2), 233–246.

Zheng, R., McAlack, M., Wilmes, B., Kohler-Evans, P., & Williamson, J. (2009). Effects of multimedia on cognitive load, self-efficacy, and multiple rule-based problem solving. *British Journal of Educational Technology*, *40*(5), 790–803.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82–91.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, *29*(3), 663–676.

Zimmerman, W. A., & Kulikowich, J. M. (2015). *Online learning self-efficacy in students with and without online learning experience.* Manuscript in preparation.

Zimmerman, W. A., & Kulikowich, J. M. (2013, October). *Development of an online learning self-efficacy scale.* Paper presented at the meeting of the Northeastern Educational Research Association, Rocky Hill, CT.

Zimmerman, W. A., Parker, D., & Knight, S. L. (2011, October). *Professional development's effect on the relationship between efficacy and behaviors in reforming Qatar.* Paper presented at the meeting of the Northeastern Educational Research Association, Rocky Hill, CT.

Appendix A: Pilot Study 1, Demonstrated Knowledge Form A

1. List two assumptions of the χ^2 test of independence.

1:

2:

2. What is the appropriate <u>null</u> hypothesis for a 2x3 χ^2 test of independence?

H_0:

3. You are conducting a χ^2 test of independence with 5 degrees of freedom and a .01 level of significance. What is your critical value?

4. In a 5x6 χ^2 test of independence, what is the numerical value of the degrees of freedom?

5. You conducted a χ^2 test of independence with two independence variables. The critical value from the χ^2 table is 5.99 and your χ^2 test statistic is 4.35. Are the two variables independent of one another? Why or why not?

6. Compute the row and column total for the following table. (i.e., fill in the empty cells)

		Sex		
		Males	Females	Row Totals
Eye Color	Blue	20	40	
	Brown	30	50	
	Other	40	20	
	Column Totals			200

7.

	Choice A	Choice B
Outcome 1	10	20
Outcome 2	40	30

Using the data in the table above, compute the expected cell values and write them in the table below. Assume that choice and outcome are independent of one another.

	Choice A	Choice B
Outcome 1		
Outcome 2		

8. Compute the χ^2 test statistic for the following table. The expected values are in parentheses.

		Coin Flip		
		Heads	Tails	Row Totals
Coin Age	New	10 (8)	8 (10)	18
	Old	6 (8)	12 (10)	18
	Column Totals	16	20	36

9. Compute the χ^2 test statistic for the following table. The expected values are in parentheses.

	Option A	Option B	Option C	Option D	Row Totals
Left-Handed	8 (10)	11 (8)	12 (10)	9 (12)	40
Right-Handed	23 (25)	22 (20)	30 (25)	25 (30)	100
Ambidextrous	19 (15)	7 (12)	8 (15)	26 (18)	60
Column Totals	50	40	50	60	200

Appendix B: Pilot Study 1, Online Data Collection Instrument Form A

PENNSTATE

COLLEGE of EDUCATION

Please read the implied inform consent form below. When you are finished, click "continue" and wait for further instructions.

Implied Informed Consent Form for Social Science Research
The Pennsylvania State University

Title of Project: Development of Self-Efficacy and Knowledge Measures (Chi-Square)

Principal Investigator: *Whitney Alicia Zimmerman, Doctoral Candidate in Educational Psychology*
Department of Educational Psychology, Counseling, and Special Education
125 CEDAR Building, University Park, PA 16802
waz107@psu.edu

Advisors: *Dr. Jonna M. Kulikowich, Professor of Education*
Department of Educational Psychology, Counseling, and Special Education
125 CEDAR Building, University Park, PA 16803
jmk35@psu.edu 814-863-2401

Dr. P. Karen Murphy, Professor of Education
Department of Educational Psychology, Counseling, and Special Education
125 CEDAR Building, University Park, PA 16803
pkm15@psu.edu 814-863-2278

1. **Purpose of the Study:** The purpose of this study is to develop instruments for measuring students' perceptions of their abilities to perform chi-square tests of independence and their abilities to demonstrate their knowledge of chi-square tests of independence. A chi-square test of independence is a statistical procedure.

2. **Procedures to be followed:** You will be asked to respond to questions concerning your confidence in your abilities to perform tasks related to online learning and chi-square tests of independence. You will be asked to rate your perceived knowledge of introductory statistics and chi-square tests of independence. You will be asked to answer questions to demonstrate your knowledge of chi-square tests of independence; some questions will be answered on the computer and some will be answered in a paper packet. You will be asked to complete a demographic questionnaire; then, you will be asked to complete the preceding instruments a second time. You will be asked for identifying information if you want to receive extra credit for your participation in this study.

3. **Duration/Time:** The estimated duration of this study is one hour. Actual completion time may vary.

4. **Statement of Confidentiality:** Your participation in this research is confidential. The questionnaire does ask for identifying information if you would like to receive extra credit for your participation. Identifying information will be removed from the dataset after instructors are given the names of their students' who participated. In the event of any publication or presentation resulting from the research, no personally identifiable information will be shared. Your confidentiality will be kept to the degree permitted by the technology being used. No guarantees can be made regarding the interception of data sent via the Internet by any third parties.

5. **Right to Ask Questions:** Please contact Whitney Alicia Zimmerman at waz107@psu.edu with questions or concerns about this study.

6. **Payment for participation:** The amount of extra credit awarded for participation in this study will be determined by your instructor. If you do not want to participate in this research study, you may complete an alternative assignment for an equal amount of extra credit. If you would like to choose this option, contact Whitney Alicia Zimmerman at waz107@psu.edu

7. **Voluntary Participation:** Your decision to be in this research is voluntary. You can stop at any time. You do not have to answer any questions you do not want to answer.

You must be 18 years of age or older to take part in this research study.

Completion and submission of the survey implies that you have read the information in this form and consent to take part in the research. Please print off this form to keep for your records.

0% 100%

Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics

PENNSTATE 1855

COLLEGE *of* EDUCATION

Please read the following directions to yourself as I read them aloud.

In this research study you are going to be asked to answer questions related to online learning and statistics.

Some of the questions will be related to the χ^2 ("Chi-Square") test of independence. You may or may not have learned about this topic in your statistics course. Please try your best and answer all questions to the best of your ability.

At the end of the study you will asked for some identifying information. This information will be given to your instructors for extra credit purposes only. Your instructors will not be given your individual responses.

You will need a pencil and calculator to answer some of the questions. If you did not bring a pencil and/or calculator, please raise your hand now and we will give you one to borrow. You may not use your cell phones.

We are going to work through each part of this research study together. Please do not work ahead or go back to previous sections.

Now, click "continue."

0% 100%

Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics

In this section you will be asked to rate a few series of statements. When you are finished, click "continue" and wait for further instructions.

Rate your confidence in your current ability to perform the following tasks related to **learning online** successfully.

A rating of 1 is the lowest; give a rating of 1 to any tasks that you have no confidence that you could complete successfully.
A rating of 9 is the highest; give a rating of 9 to any tasks that you have complete confidence that you could complete successfully.

	No Confidence							Complete Confidence	
	1	2	3	4	5	6	7	8	9
Submit assignments to an online dropbox	○	○	○	○	○	○	○	○	○
Overcome technical difficulties on my own	○	○	○	○	○	○	○	○	○
Learn without being in the same room as other students	○	○	○	○	○	○	○	○	○
Develop and follow a plan for completing all required work on time	○	○	○	○	○	○	○	○	○
Focus on schoolwork when faced with distractions	○	○	○	○	○	○	○	○	○
Meet deadlines with very few reminders	○	○	○	○	○	○	○	○	○
When a problem arises, promptly ask questions in the appropriate forum (email, discussion board, etc.)	○	○	○	○	○	○	○	○	○
Complete all assignments on time	○	○	○	○	○	○	○	○	○
Communicate effectively with technical support via email, telephone, or live online chat	○	○	○	○	○	○	○	○	○
Communicate using asynchronous technologies (discussion boards, email, etc.)	○	○	○	○	○	○	○	○	○
Use the library's online resources efficiently	○	○	○	○	○	○	○	○	○
Navigate the online grade book	○	○	○	○	○	○	○	○	○
	1	2	3	4	5	6	7	8	9
Use synchronous technology to communicate with others (such as Skype)	○	○	○	○	○	○	○	○	○
Navigate online course materials efficiently	○	○	○	○	○	○	○	○	○
Learn from videos	○	○	○	○	○	○	○	○	○
Communicate effectively with my instructor via email	○	○	○	○	○	○	○	○	○
Learn to use a new type of technology efficiently	○	○	○	○	○	○	○	○	○
Learn without being in the same room as the instructor	○	○	○	○	○	○	○	○	○
Search the Internet to find the answer to a course-related question	○	○	○	○	○	○	○	○	○
Complete a group project entirely online	○	○	○	○	○	○	○	○	○
Find the course syllabus online	○	○	○	○	○	○	○	○	○
Manage time effectively	○	○	○	○	○	○	○	○	○
Search the online course materials	○	○	○	○	○	○	○	○	○
Complete an online statistics course with a grade of A or B	○	○	○	○	○	○	○	○	○
	1	2	3	4	5	6	7	8	9

Rate your confidence in your current ability to perform the following tasks related to **performing statistics** successfully.

A rating of 1 is the lowest; give a rating of 1 to any tasks that you have no confidence that you could complete successfully.
A rating of 9 is the highest; give a rating of 9 to any tasks that you have complete confidence that you could complete successfully.

	No Confidence								Complete Confidence
	1	2	3	4	5	6	7	8	9
Identify the assumptions of a χ^2 test of independence	○	○	○	○	○	○	○	○	○
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	○	○	○	○	○	○	○	○	○
Look up the appropriate critical value on a χ^2 table	○	○	○	○	○	○	○	○	○
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	○	○	○	○	○	○	○	○	○
Calculate row and column totals	○	○	○	○	○	○	○	○	○
Calculate expected cell values	○	○	○	○	○	○	○	○	○
Compute the test statistic for a 2x2 χ^2 test of independence	○	○	○	○	○	○	○	○	○
Compute the test statistic for a 3x4 χ^2 test of independence	○	○	○	○	○	○	○	○	○
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	○	○	○	○	○	○	○	○	○

How much do you know about the following topics?

A rating of 1 is the lowest; give a rating of 1 to any topic that you know nothing about.
A rating of 9 is the highest; give a rating of 9 to answer topic that you know a great deal about.

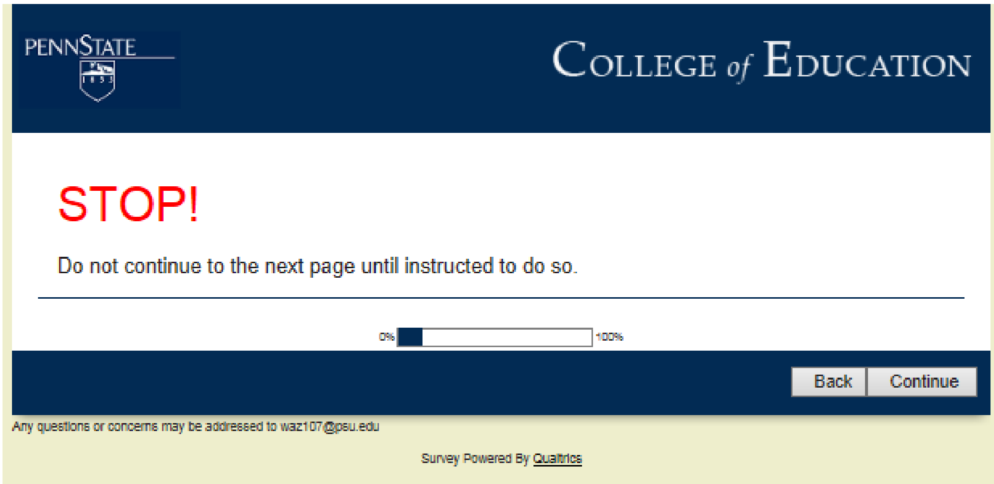
	Nothing								A Great Deal
	1	2	3	4	5	6	7	8	9
Introductory statistics	○	○	○	○	○	○	○	○	○
χ^2 Tests of Independence	○	○	○	○	○	○	○	○	○

When you are finished answering the questions on this page, click "continue" and wait for further instructions.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu



PENNSTATE

COLLEGE of EDUCATION

STOP!

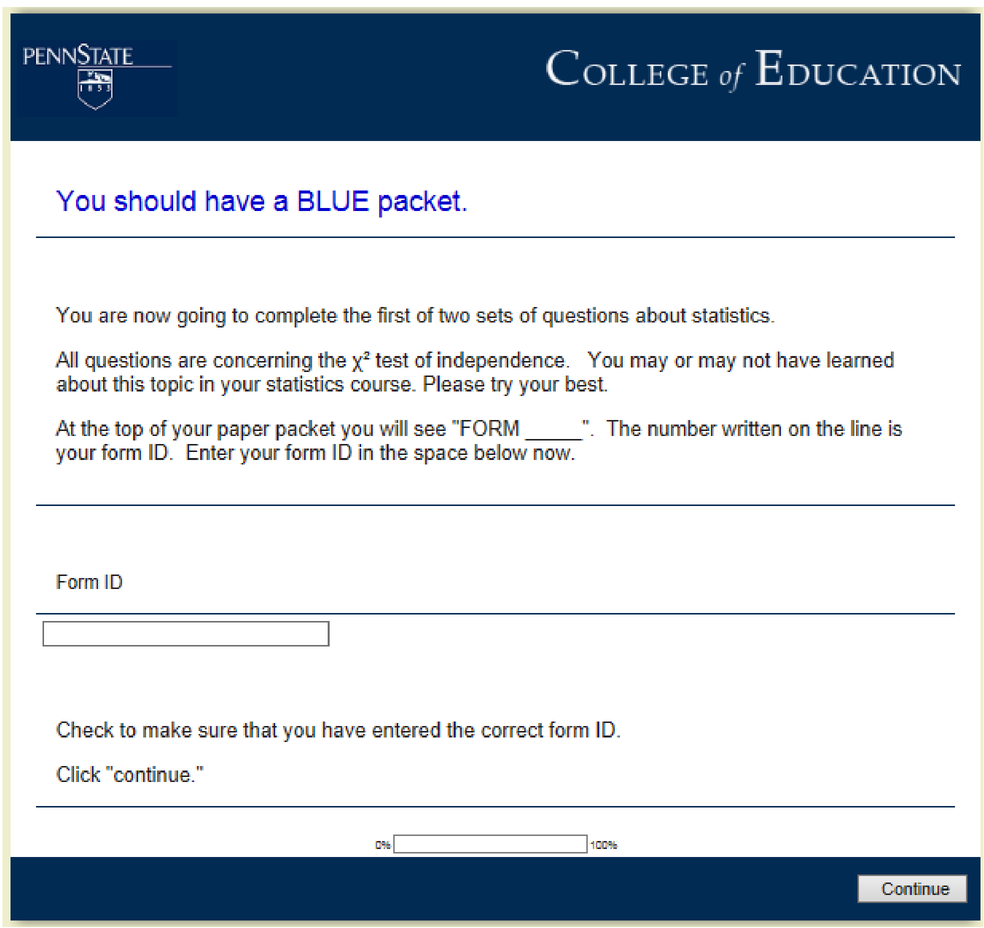
Do not continue to the next page until instructed to do so.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics



PENNSTATE

COLLEGE of EDUCATION

You should have a BLUE packet.

You are now going to complete the first of two sets of questions about statistics.

All questions are concerning the χ^2 test of independence. You may or may not have learned about this topic in your statistics course. Please try your best.

At the top of your paper packet you will see "FORM _____". The number written on the line is your form ID. Enter your form ID in the space below now.

Form ID

Check to make sure that you have entered the correct form ID.

Click "continue."

0% 100%

Continue

This is the first of two sets of questions about statistics.

There are a total of 9 questions. There are 5 questions that you will answer on the computer and 4 questions that you will answer in your packet. After you open your packet, these questions are on the first two pages of your packet. Only complete these two pages now (i.e., page 2 and page 3). You will complete the last two pages of this packet at the end of the study.

Formulas and a table that you may need are given below. After you have completed these 9 questions, click "continue" and wait for further instructions. Do not click "continue" until you have answered all 9 questions, including those in your packet!

$$\chi^2 = \sum \left[\frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}} \right]$$

Where $\mathrm{O} = \text{Observed cell frequency}$

$\mathrm{E} = \text{Expected cell frequency}$

$$\mathrm{E} = \frac{\text{Row total } \times \text{Column total}}{\text{Total Sample Size}}$$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

1. List two assumptions of the χ^2 test of independence

1:

2:

2. What is the appropriate <u>null</u> hypothesis for a 2x3 χ^2 test of independence?

H_0:

3. You are conducting a χ^2 test of independence with 5 degrees of freedom and a .01 level of significance. What is your critical value?

4. In a 5x6 χ^2 test of independence, what is the numerical value of the degrees of freedom?

5. You conducted a χ^2 test of independence with two independent variables. The critical value from the χ^2 table is 5.99 and your χ^2 test statistic is 4.35. What is your conclusion concerning the independence of the two variables? Why?

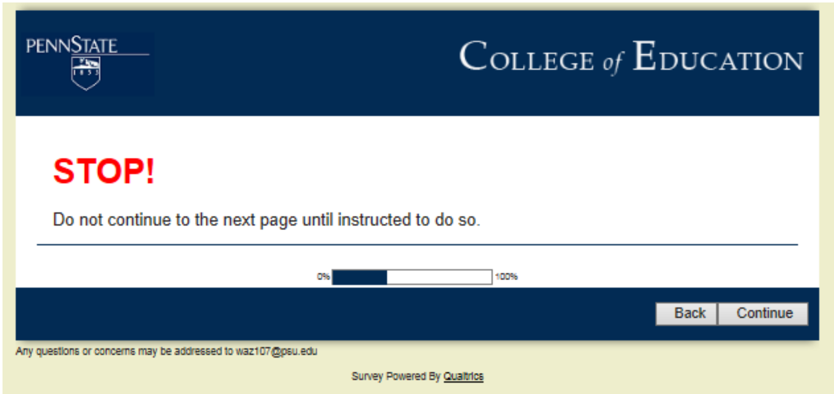
Questions 6 through 9 you will complete in your packet. Show all of your work. The formulas and table that you may need are given above.

After you have answered all 9 questions, close your packet, click "continue," and wait for further instructions. Do not click continue until you have answered all 9 questions.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu



PENNSTATE

COLLEGE of EDUCATION

STOP!

Do not continue to the next page until instructed to do so.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics

Please respond to the following demographic questions.

You do not need to answer any questions that you feel uncomfortable with.

When you are finished, click "continue" and wait for further instructions.

What is your age?

What is your Gender?

Female
Male
Other

What is your academic major?

What is your primary campus?

- ◯ University Park
- ◯ World Campus
- ◯ Other [ ]

What is your current status?

- Freshman
- Sophomore
- Junior
- Senior
- Graduate Student
- Non-degree
- Other

What is your GPA at Penn State? If a first semester student, leave blank.

[ ]

How many statistics courses have you completed? Do not count those that you are currently enrolled in. Enter a whole number. If none, enter "0".

[ ]

Have you learned about χ^2 tests of independence in a statistics course? (Either in the past or in a course that you are currently enrolled in)

- ◯ Yes
- ◯ No

Are you currently enrolled in an online statistics course?

- ○ Yes
- ○ No

How many for-credit online courses have you completed? Do not count those that you are currently enrolled in. Enter a whole number. If none, enter "0".

How many for-credit online courses are you currently enrolled in? Enter a whole number. If none, enter "0".

Do you have any experience with online education other than for-credit online courses? If yes, please describe (e.g., online teaching experience, MOOCs)

When you are finished answering the questions on this page, click "continue" and wait for further instructions.

0% 100%

Back Continue

ny questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics

PENNSTATE

COLLEGE of EDUCATION

STOP!

Do not continue to the next page until instructed to do so.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics

In this section you will be asked to rate a few series of statements. These are the same statements that you rated earlier.

When you are finished, click "continue" and wait for further instructions.

Rate your confidence in your current ability to perform the following tasks related to **learning online** successfully.

A rating of 1 is the lowest; give a rating of 1 to any tasks that you have no confidence that you could complete successfully.
A rating of 9 is the highest; give a rating of 9 to any tasks that you have complete confidence that you could complete successfully.

	No Confidence								Complete Confidence
	1	2	3	4	5	6	7	8	9
Complete an online statistics course with a grade of A or B	○	○	○	○	○	○	○	○	○
Use the library's online resources efficiently	○	○	○	○	○	○	○	○	○
Complete a group project entirely online	○	○	○	○	○	○	○	○	○
Find the course syllabus online	○	○	○	○	○	○	○	○	○
Use synchronous technology to communicate with others (such as Skype)	○	○	○	○	○	○	○	○	○
Learn without being in the same room as the instructor	○	○	○	○	○	○	○	○	○
Communicate effectively with my instructor via email	○	○	○	○	○	○	○	○	○
Communicate effectively with technical support via email, telephone, or live online chat	○	○	○	○	○	○	○	○	○
Navigate the online grade book	○	○	○	○	○	○	○	○	○
When a problem arises, promptly ask questions in the appropriate forum (email, discussion board, etc.)	○	○	○	○	○	○	○	○	○
Search the Internet to find the answer to a course-related question	○	○	○	○	○	○	○	○	○
Communicate using asynchronous technologies (discussion boards, email, etc.)	○	○	○	○	○	○	○	○	○
	1	2	3	4	5	6	7	8	9
Learn to use a new type of technology efficiently	○	○	○	○	○	○	○	○	○
Learn without being in the same room as other students	○	○	○	○	○	○	○	○	○
Search the online course materials	○	○	○	○	○	○	○	○	○
Complete all assignments on time	○	○	○	○	○	○	○	○	○
Navigate online course materials efficiently	○	○	○	○	○	○	○	○	○
Learn from videos	○	○	○	○	○	○	○	○	○
Develop and follow a plan for completing all required work on time	○	○	○	○	○	○	○	○	○
Manage time effectively	○	○	○	○	○	○	○	○	○
Focus on schoolwork when faced with distractions	○	○	○	○	○	○	○	○	○
Meet deadlines with very few reminders	○	○	○	○	○	○	○	○	○
Overcome technical difficulties on my own	○	○	○	○	○	○	○	○	○
Submit assignments to an online dropbox	○	○	○	○	○	○	○	○	○
	1	2	3	4	5	6	7	8	9

Rate your confidence in your current ability to perform the following tasks related to **statistics** successfully.

A rating of 1 is the lowest; give a rating of 1 to any tasks that you have no confidence that you could complete successfully.
A rating of 9 is the highest; give a rating of 9 to any tasks that you have complete confidence that you could complete successfully.

	No Confidence								Complete Confidence
	1	2	3	4	5	6	7	8	9
Identify the assumptions of a χ^2 test of independence	○	○	○	○	○	○	○	○	○
Write the appropriate null and alternative hypotheses for a χ^2 test of independence	○	○	○	○	○	○	○	○	○
Look up the appropriate critical value on a χ^2 table	○	○	○	○	○	○	○	○	○
Calculate the appropriate degrees of freedom for a given χ^2 test of independence	○	○	○	○	○	○	○	○	○
Calculate row and column totals	○	○	○	○	○	○	○	○	○
Calculate expected cell values	○	○	○	○	○	○	○	○	○
Compute the test statistic for a 2x2 χ^2 test of independence	○	○	○	○	○	○	○	○	○
Compute the test statistic for a 3x4 χ^2 test of independence	○	○	○	○	○	○	○	○	○
Given a χ^2 test statistic, make the correct decision whether to reject or fail to reject the null hypothesis	○	○	○	○	○	○	○	○	○

How much do you know about the following topics?

A rating of 1 is the lowest; give a rating of 1 to any topic that you know nothing about.
A rating of 9 is the highest; give a rating of 9 to answer topic that you know a great deal about.

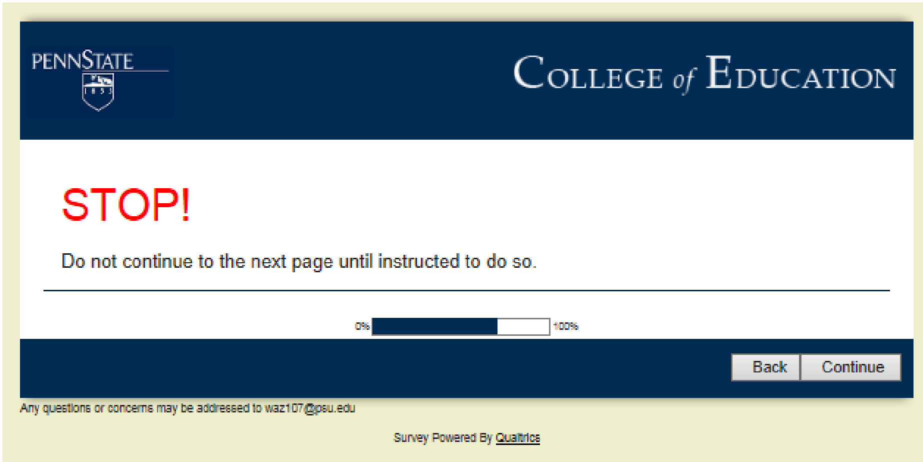
	Nothing								A Great Deal
	1	2	3	4	5	6	7	8	9
Introductory Statistics	○	○	○	○	○	○	○	○	○
χ^2 Tests of Independence	○	○	○	○	○	○	○	○	○

When you are finished answering the questions on this page, click "continue" and wait for further instructions.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu



You are now going to complete the second set of questions about statistics.

There are a total of 9 questions. There are 5 questions that you will answer on the computer and 4 questions that you will answer in your packet. These questions are on the last two pages of your packet. Work only on the last two pages. Do not return to the previous section.

Formulas and a table that you may need are given below. After you have completed all 9 questions, click "continue" and wait for further instructions. Do not click continue until you have answered all 9 questions, including those in your packet.

$$\chi^2 = \sum \left[\frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}} \right]$$

Where $\mathrm{O} = \text{Observed cell frequency}$

$\mathrm{E} = \text{Expected cell frequency}$

$$\mathrm{E} = \frac{\text{Row total} \times \text{Column total}}{\text{Total Sample Size}}$$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

1. List two assumptions of the χ^2 test of independence

1:

2:

2. What is the appropriate <u>alternative</u> hypothesis for a 2x2 χ^2 test of independence?

H_A:

3. You are conducting a χ^2 test of independence with 11 degrees of freedom and a .05 level of significance. What is your critical value?

4. In a 3x7 χ^2 test of independence, what is the numerical value of the degrees of freedom?

5. **You conducted a χ^2 test of independence with two independent variables. The critical value from the χ^2 table is 7.82 and your χ^2 test statistic is 28.25. What is your conclusion concerning the independence of the two variables? Why?**

Questions 6 through 9 you will complete in your packet. Show all of your work. The formulas and table that you may need are given above.

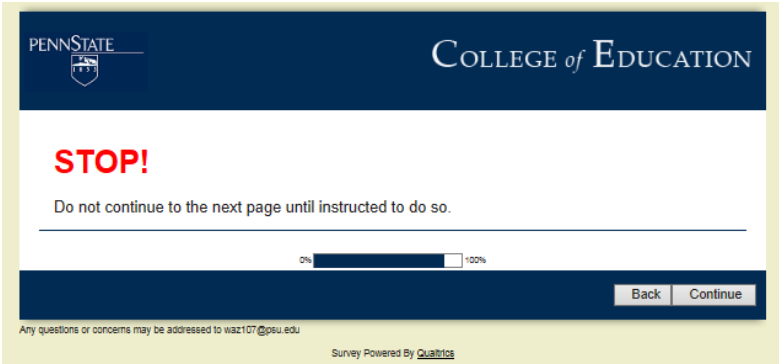
After you have answered all 9 questions, click "continue" and wait for further instructions. Do not click continue until you have answered all 9 questions.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics



PENNSTATE

COLLEGE of EDUCATION

STOP!

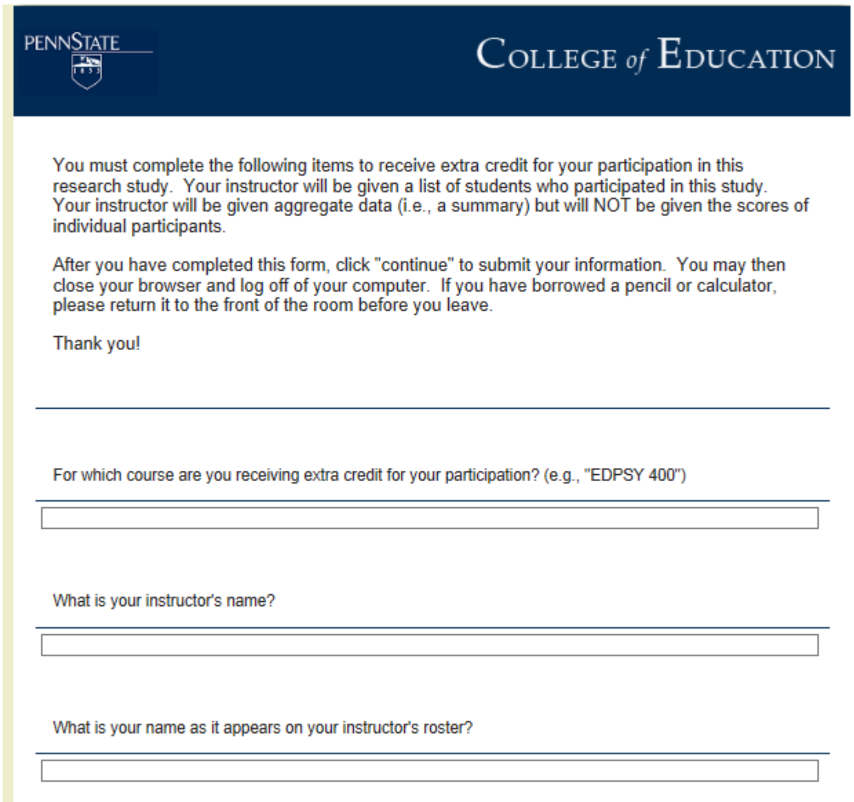
Do not continue to the next page until instructed to do so.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics



PENNSTATE

COLLEGE of EDUCATION

You must complete the following items to receive extra credit for your participation in this research study. Your instructor will be given a list of students who participated in this study. Your instructor will be given aggregate data (i.e., a summary) but will NOT be given the scores of individual participants.

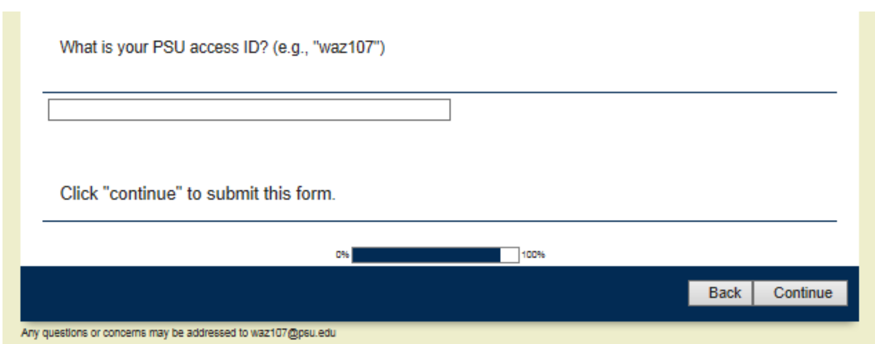
After you have completed this form, click "continue" to submit your information. You may then close your browser and log off of your computer. If you have borrowed a pencil or calculator, please return it to the front of the room before you leave.

Thank you!

For which course are you receiving extra credit for your participation? (e.g., "EDPSY 400")

What is your instructor's name?

What is your name as it appears on your instructor's roster?



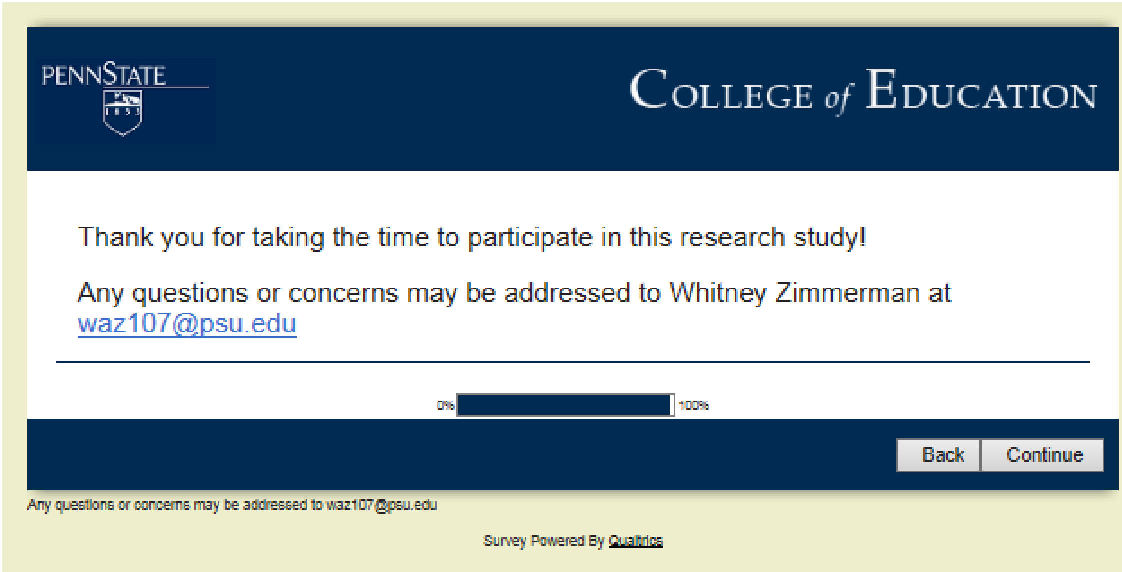
What is your PSU access ID? (e.g., "waz107")

Click "continue" to submit this form.

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu



PENNSTATE

COLLEGE *of* EDUCATION

Thank you for taking the time to participate in this research study!

Any questions or concerns may be addressed to Whitney Zimmerman at waz107@psu.edu

0% 100%

Back Continue

Any questions or concerns may be addressed to waz107@psu.edu

Survey Powered By Qualtrics

Appendix C: Pilot Study 2, Video Presentation Slides

χ^2 Test of Independence

χ^2 Test of Independence

- Used to determine if two categorical variables are related
- Compares observed frequencies to expected frequencies

Example

Research Question: Are dog ownership and cat ownership related?

		Dog Owner	
		Yes	No
Cat Owner	Yes	10	20
	No	15	30

Assumptions

- Random sample
- Categorical variables
- Each cell must contain an expected value count of at least 5
- Each individual appears in only one cell

Hypotheses

$\mathrm{H_0}$: Dog and cat ownership are independent

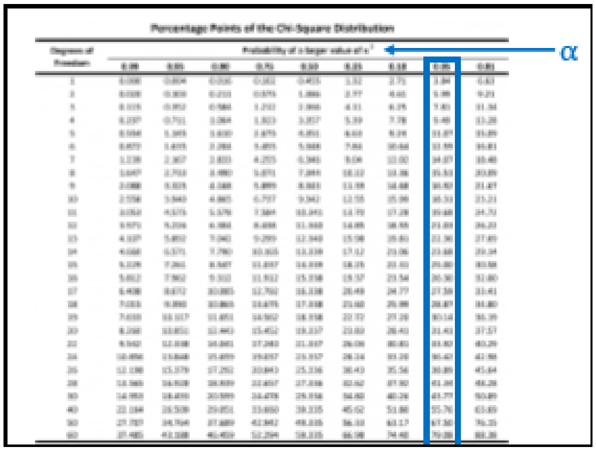
$\mathrm{H_a}$: Dog and cat ownership are not indpendent

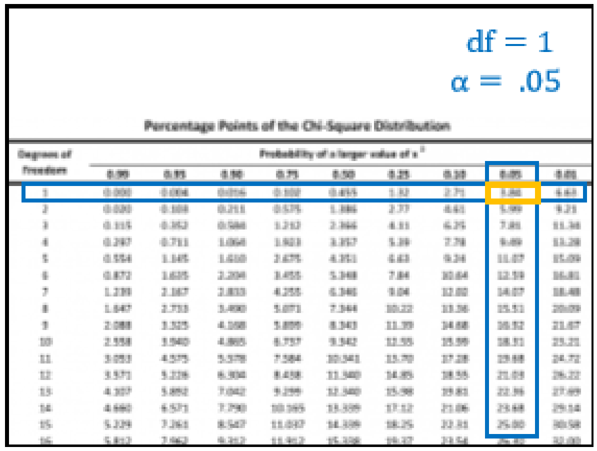
χ^2 Distribution

$$\mathrm{df} = (\mathrm{Rows} - 1)(\mathrm{Columns} - 1)$$

		Dog Owner	
		Yes	No
Cat Owner	Yes	10	20
	No	15	30

$$\mathrm{df} = (2 - 1)(2 - 1) = (1)(1) = 1$$





Computing a χ^2 Test Statistic

1. Compute row and column totals
2. Compute expected values
3. Compute the test statistic

1. Compute Row and Column Totals

		Dog Owner		
		Yes	No	
Cat Owner	Yes	30	25	55
	No	10	35	45
		40	60	100

2. Compute expected values

$$\text{Expected} = \frac{(\text{Row Total})(\text{Column Total})}{\text{Grand Total}}$$

		Dog Owner		
		Yes	No	
Cat Owner	Yes	$\frac{55*40}{100}=22$	$\frac{55*60}{100}=33$	55
	No	$\frac{45*40}{100}=18$	$\frac{45*60}{100}=27$	45
		40	60	100

3. Compute the test statistic

$$\chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$$

		Dog Owner	
		Yes	No
Cat Owner	Yes	30 (22)	25 (33)
	No	10 (18)	35 (27)

$$\chi^2 = \frac{(30-22)^2}{22} + \frac{(25-33)^2}{33} + \frac{(10-18)^2}{18} + \frac{(35-27)^2}{27}$$

$$\chi^2 = 2.909 + 1.939 + 3.556 + 2.370$$

$$\chi^2 = 10.774$$

Research question: Is there a relationship between hair color and whether or not someone drinks coffee?

H_0: Dog and cat ownership are independent
H_a: Dog and cat owndership are related

df = 1, α = .05 $\chi^2_{CV} = 3.84$

$\chi^2_{stat} = 10.774$

Reject the null hypothesis. There is evidence of a relationship between dog and cat ownership.

Summary

- The χ^2 test of independence is used to determine if two categorical variables are related
- Compares observed and expected frequencies
- Assumes a random sample
- All expected values must be ≥ 5
- Each individual can only appear in one cell
- To compute the χ^2 test statistic:
  - Compute row and column totals
  - Compute expected values
  - Compute χ^2 test statistic

Appendix D: Pilot Study 2, Video Presentation Script

Slide 1 (Title):
This presentation will give a general introduction to the chi-square test of independence and will walk you through one example.

Slide 2 (Intro):
*The chi-square test of independence is a statistical analysis that is used to determine if categorical variables are related.
*It compares observed frequencies to the expected frequencies. In other words, it compares the values obtained in your sample to the values that would be expected if the two variables were independent of one another.

Slide 3 (Example):
Let's look at an example.
* Consider the following research question. Are dog ownership and cat ownership related?
* We surveyed 100 people and asked them if they own a dog and/or a cat. We used that data to make what we call a contingency table. On this contingency table, each person fits into one and only one cell. For example, 30 people said that they own both a dog and a cat. 25 people said that they own a cat but not a dog, and so on. We're going to conduct a chi-square test of independence using this data in a moment, but first let's assumptions and hypotheses of the test.

Slide 4 (Assumptions):
There are a number of assumptions of the chi-square test of independence. The test is only appropriate when all assumptions have been met.
* As with most inferential statistics procedures, a random sample is required. A random sample is drawn from a population in a manner such that every member of the population has an equal chance of being selected.
* The variables must be categorical in nature. In other words, they must be on a nominal or ordinal level. If your data are on an interval or ratio level scale they must be converted to categories.
* Each cell must contain an expected value count of at least 5. When we work through the example we will compute the expected cell counts; each of these must be a minimum of 5.
* Each individual appears in only one cell. In our example, each person had to answer yes or no to each question. They had to answer both questions in order to fit into one cell. Now let's look at the hypotheses of the test.

Slide 5 (Hypotheses):
The null hypothesis, denoted as H subscript 0, is that the variables are independent. In this example, the null hypothesis is that dog and cat ownership are independent. When we say that two variables are independent that means that they are not related to one another. The alternative hypothesis, denoted as H subscript a, is that the variables are not independent. Here, our alternative hypothesis is that dog and cat ownership are no independent. In other words, they are related.

Slide 6 (χ^2 Distribution):
The chi-square distribution is used in way similar to that of the z distribution or t distribution. Like the t distribution, the chi-square distribution takes into account degrees of freedom.

* For the chi-squared test of independence, the degrees of freedom are equal to the number of rows minus one times the number of columns minus one.
*In our example, there are two rows and two columns.
* Therefore the degrees of freedom are two minus one times two minus one, or, one times one which equals one. There is one degree of freedom in this example.

Slide 7 (χ^2 Table):
This is the chi-square table that is used to look up critical values. To find a critical value we need two pieces of information: the degrees of freedom and the alpha value. On this table, the degrees of freedom are presented vertically. The alpha values are labeled “Probability of a larger value of χ^2” and are presented horizontally. The most common alpha value in social science research is .05. Let’s go back and look at our example again.

Slide 8 (Critical Value Example):
Our example had one degree of freedom and an alpha level of .05
* We move down to one degree of freedom.
* Then, we move over to our alpha level of .05
* Our chi-square critical value is 3.84. This critical value will be compared to the test statistic which we will compute next. If our test statistic is greater than 3.84 we will reject the null hypothesis. If our test statistic is less than or equal to 3.84 we will fail to reject the null hypothesis. Now let’s compute our test statistic.

Slide 9 (Compute a χ^2 Test Statistic):
We are going to complete the following three steps to compute a chi-square test statistic. Using our example data, first we’ll compute row and column totals. Second, we’ll compute expected value. And finally, we’ll compute the test statistic.

Slide 10 (Compute Row and Column Totals):
Step one is to compute the row and column totals.
* Let’s look at our example data. We can start by computing the first row’s total. This will be the total number of people who said ‘yes’ they own a cat. This is 30 plus 25 which equals 55.
* Now we’ll do the same for the second row; this is the total number of people who said ‘no’ they do not own a cat. 10 plus 35 equals 45.
* We also need to compute the column totals. The number of participants who own a dog is 30 plus 10 which equals 40.
* The number of participants who do not own a dog is 25 plus 35 which equals 60.
* These are our row and column totals. Let’s also compute the total number of participants, or the grand total. To compute this we can add all of the cell frequencies. 30 plus 25 plus 10 plus 35 equals 100.
* Now that we’ve computed all of the row totals, columns totals, and the grand total, we can move to step 2.

Slide 11 (Compute Expected Values):
To compute the expected value for each cell we use the formula expected equals the product of row total and column total divided by the grand total.
* Let’s look at our example with the row, column, and grand totals. We can start by computing the expected value for the first cell. These are the individuals who identified as owning both a dog and a cat. The row total associated with owning a cat is 55; the column total associated with owning a dog is

40. The grand total number of participants is 100. Thus, the expected cell value will be 55 times 40 divided by 100 which equals 22. If dog and cat ownership are independent, we would expect about 22 people in our sample to own both a dog and a cat.
*We'll do the same for the remaining three cells. For individuals who own a cat but not a dog, 55 times 60 divided by 100 equals 33.
* For individuals who own a dog but not a cat, 45 times 40 divided by 100 equals 18.
* Finally, for individuals who do not own a dog and do not own a cat, 45 times 60 divided by 100 equals 27.
* These are all of our expected values. The expected values take into account the row and column totals. If the two variables are independent, these are what we would expect the cell values to be. Now, we can use these values to compute the test statistic.

Slide 12 (Compute the Test Statistic):
The chi-square test statistic is computed using this formula. Here, the O is the observed values; these are the actual cell frequencies that we observed in our sample. The E is the expected values that we calculated in the last step. Let's look at our example again.
* Here, in each cell we have the observed value first, and second we have the expected value in parentheses. For example, for individuals who own both a dog and a cat, we observed 30 people in our sample and we computed an expected value of 22. Now let's apply the formula to compute the chi-square test statistic.
* The first fraction here is for the first cell. The observed value is 30, minus the expected value of 22. This is squared and then divided by the expected value of 22. We follow this pattern for each of the four cells. Take a moment to study this pattern. [Pause]
* * This then simplifies to 10.774. Our chi-square test statistic is 10.774. We will compare this value to the critical value that we looked up on the table. But first, let's review our research question.

Slide 13 (Hypothesis Testing Steps):
Our original research question was, is there a relationship between dog and cat ownership?
* Our null hypothesis was that dog and cat ownership are independent, in other words not related. Our alternative hypothesis was that dog and cat ownership are related, in other words, not independent.
*Our test had one degree of freedom and an alpha level of .05. We used the chi-square table to look up a critical value of 3.84.
*Then, we computed a chi-square test statistic of 10.774 by using our observed and expected values.
*Our test statistic of 10.774 is greater than our critical value of 3.84; therefore we should reject the null hypothesis. There is evidence of a relationship between dog and cat ownership. This is our final conclusion. Our results were statistically significant in this example. Let's review some of the main points that we have covered here.

Slide 14 (Summary):
* The chi-square test of independence is used to determine if categorical variables are related. Today, we only considered two variables at a time, but it is possible to work with three or more variables.
* The chi-square test of independence compares observed and expected frequencies. The observed frequencies are the counts that were obtained from our sample. The expected frequencies we computed; these are the cell values that we would expected if the variables were really independent.
* The test assumes a random sample.
* All expected values must be greater than or equal to 5.
* Each individual can only appear in one cell.
* And, to compute the chi-square test statistic we follow three steps:

* First, compute row and column totals
* Second, compute expected values
* Third, compute the chi-square test statistic. This concludes this video presentation concerning the chi-square test of independence. Next, you will be given practice problems to work through.

Appendix E: Pilot Study 2, Conventional Practice Problems

Problem #1

You are conducting a research study on beverage preferences of preschoolers. You want to know if there is a relationship between gender and whether children prefer milk or water. Use the following data to answer this research question. (α = .05)

		Preferred Beverage	
		Milk	Water
Gender	Boy	17	13
	Girl	23	12

Problem #2

Consider the following research question: Is there a relationship between whether someone is from the North or South and whether they prefer Coke or Pepsi? (α = .10)

		Region	
		North	South
Cola Preference	Coke	30	20
	Pepsi	15	40

Problem #3

You want to know if there is a relationship between student level (undergraduate versus graduate) and whether they get more or less than eight hours of sleep on an average night. Use the following data to answer this research question. (α = .05)

		Average Nightly Sleep	
		Less than 8 hours	**More than 8 hours**
Student Level	**Undergraduate**	**20**	**20**
	Graduate	**15**	**10**

Appendix F: Pilot Study 2, Worked Example Practice Problem

Problem #1: Example

You are conducting a research study on beverage preferences of preschoolers. You want to know if there is a relationship between gender and whether children prefer milk or water. Use the following data to answer this research question. (α = .05)

		Preferred Beverage	
		Milk	**Water**
Gender	**Boy**	**17**	**13**
	Girl	**23**	**12**

H_0: Gender and beverage preference are independent

H_a: Gender and beverage preference are not indpendent

Find the degrees of freedom and critical value

$$df = (\text{Rows} - 1)(\text{Columns} - 1) = (2 - 1)(2 - 1) = 1$$

Degrees of Freedom	Probability of a larger value of x [2]								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34

$\chi^2_{CV} = 3.84$

Compute row and column totals

		Preferred Beverage		
		Milk	Water	
Gender	Boy	17	13	30
	Girl	23	12	35
		40	25	65

Compute expected values

$$\mathrm{E} = \frac{\text{Row total} \times \text{Column total}}{\text{Total Sample Size}}$$

$$E_{Boy,Milk} = \frac{30*40}{65} = 18.462 \qquad E_{Boy,Water} = \frac{30*25}{65} = 11.538$$

$$E_{Girl,Milk} = \frac{35*40}{65} = 21.538 \qquad E_{Girl,Water} = \frac{35*25}{65} = 13.462$$

Compute the test statistic

		Preferred Beverage		
		Milk	Water	
Gender	Boy	17 (18.462)	13 (11.538)	30
	Girl	23 (21.538)	12 (13.462)	35
		40	25	65

$$\chi^2 = \sum \left[\frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}} \right]$$

$$\chi^2 = \frac{(17-18.462)^2}{18.462} + \frac{(13-11.538)^2}{11.538} + \frac{(23-21.538)^2}{21.538} + \frac{(12-13.462)^2}{13.462}$$

$$= 0.116 + 0.185 + 0.099 + 0.159$$

$$= 0.559$$

State your conclusion

The test statistic (0.559) is less than the critical value (3.84); fail to reject the null hypothesis.

There is not evidence that gender and beverage preference are related.

Problem #2 and Problem #3 were the same for the worked example and conventional problem solving groups.

Abbreviated Vita

Whitney Alicia Zimmerman

Education

The Pennsylvania State University, University Park, Pennsylvania
M.S., Educational Psychology, 2011
M.Ed., Adult Education, 2012
Graduate Certificate, Applied Statistics, 2012
Post-Baccalaureate Certificate, Distance Education, 2011
Susquehanna University, Selinsgrove, Pennsylvania
B.S., Psychology, 2007

Publications

Shattuck, K. I., Zimmerman, W. A., & Adair, D. (2014). Continuous improvement of the QM Rubric and review processes: Scholarship of integration and application. *Internet Learning, 3*(1).

Knight, S. L., Parker, D., Zimmerman, W. A., & Ikhlef, A. (2014). Investigating the relationship between perceived and observed student-centered learning environments in Qatari elementary math and science classrooms. *Learning Environments Research: An International Journal*, *17*(1), 29-47. DOI: 10.1007/s10984-014-9156-z

Select Conference Presentations

Zimmerman, W. A., & Goins, D. D. (2014, October). Impact of experiences on calibration between self-efficacy and demonstrated knowledge in introductory statistics. Paper presented at the Northeastern Educational Research Association's 2014 Annual Meeting, Trumbull, CT.

Zimmerman, W. A., & Kulikowich, J. M. (2013, October). *Development of an online learning self-efficacy scale.* Paper presented at the Northeastern Educational Research Association's 2013 Annual Meeting, Rocky Hill, CT.

Zimmerman, W. A., Kulikowich, J. M., Corrigan, S., & Bravo, M. A. (2012, October). *Impact of English language learner status on the measurement of science learning-related constructs.* Paper presented at the Northeastern Educational Research Association's 2012 Annual Meeting, Rocky Hill, CT.

Zimmerman, W. A. (2012, October). *Revisiting Parten and Piaget: Gender differences in the social and cognitive play of young children.* Poster presented at the Northeastern Educational Research Association's 2012 Annual Meeting, Rocky Hill, CT.

Corrigan, S., & Zimmerman, W. A. (2012, April). *Measuring the impact of ECMs on teacher knowledge/efficacy and ELL student learning.* Paper presented at the meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Knight, S. L., Parker, D., Zimmerman, W., & Ikhlef, A. (2012, April). *Investigating the relationship between professional development and student-centered learning environment in Qatari math and science elementary classrooms.* Paper presented at the meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Zimmerman, W. A. (2010, June). *Quality Matters 2008-2010 rubric analyses.* Invited presentation at the 2nd Annual Quality Matters Conference, Oakbrook, IL.

Sperling, R., Kulikowich, J., Ramsay, C., Betaudier, T., Zimmerman, W., & Higley, K. (2010, May). *Determining the alignment to standards and the validity of district-level assessment practices.* Paper at the meeting of the American Educational Research Association, Denver, CO.