A Model for Improving Survey Outcomes by Reducing Cognitive Load

BY

JENNIFER SWEET
B.S., University of Wisconsin – River Falls, 2002
M.S., Oklahoma State University, Stillwater, 2005

DISSERTATION

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Chicago, 2016

Chicago, Illinois

Defense Committee:

Carol Myford, Chair and Advisor
Ranfen Li, Academic Program Review and Assessment
Carol-Ann Emmons, Illinois Institute of Technology
Yue Yin, Educational Psychology
Everett Smith, Educational Psychology

ACKNOWLEDGEMENTS

First, I would like to acknowledge my advisor, Dr. Carol Myford. Without Carol's knowledge, support, and guidance, I would have never been able to complete my dissertation. Throughout the dissertation process, Carol was unwaveringly warm, supportive, and understanding. I did not have an easy or typical path to completing my dissertation, and her support and unique understanding of my situation was helpful in ways I cannot fully describe.

I would also like to thank Dr. Ranfen Li, my former boss and current colleague, friend, and committee member. Ranfen introduced me to the field of higher education assessment and has always been incredibly supportive of my professional development, including my pursuit of my doctoral degree.

I would like to acknowledge the additional members of my committee, Dr. Carol-Ann Emmons, Dr. Everett Smith, and Dr. Yue Yin whose feedback contributed greatly to this research. Without their feedback and support, I would not have been able to complete this dissertation.

My family also contributed to the completion of this dissertation. My family has supported me in every way they can--largely by helping me to maintain my emotional health, but also by taking me in two weeks every summer to let me focus exclusively on my dissertation work.

Finally, I would like to thank God. About halfway into this process, I was diagnosed with an aggressive and terminal form of brain cancer. Based on my initial diagnosis, I should not have lived more than a few months. However, my faith in God helped me to never give up on my life, including the pursuit of my PhD, and God's grace has allowed me to live and flourish far beyond the time medical science would have predicted.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

LIST OF FIGURES

FIGURE PAGE

LIST OF ABBREVIATIONS

ERS	Extreme Response Styles
HCL	High Cognitive Load
LCL	Low Cognitive Load

SUMMARY

Surveys are a very common method by which people collect information. However, most people who design surveys are not experts in survey research. Those who seek to design surveys based on the myriad recommendations and best practices in the wide-ranging survey design literature face a daunting task. Currently, there is no theoretical model that practitioners can use to design surveys. In this research, I conducted two studies to determine the utility of using the concept of *cognitive load* (the amount of mental effort or "thinking" required to respond to survey items) to construct a model that practitioners could employ to develop surveys.

The purposes of the studies were to (1) create a succinct and easy-to-understand model for reducing cognitive load in survey items and instruments, and then (2) test the efficacy and utility of a portion of that model. The model is based on key findings from research on survey design and cognitive load research. I used the model to design online survey instruments to measure the level of agreement that students indicated when they responded to a series of statements regarding satisfaction with their college experience.

For my first study, I created a survey instrument that presented students (n = 64) with two versions of each item. The two versions measured the same content but differed in their cognitive load. One was a high cognitive load (HCL) version, and the other was a low cognitive load (LCL) version. Students read the two versions of each item and then decided which version required more mental effort to respond. For each item, I used a z-test to compare the proportion of students who selected the HCL version to the proportion expected if students were to respond randomly when making their selections.

In a second study, I randomly assigned students to complete a survey that contained all the HCL versions of the items (n = 280 students) or a survey that contained all the LCL versions

of the items (n = 277 students). I calculated the response rate for each survey and for each item on a survey, the time students took to respond to each of the surveys, and Rasch student fit statistics and point-measure correlations to detect response sets in the students' ratings.

Overall, students reported that the HCL items required more mental effort to respond to than the LCL items, and they spent significantly more time responding to items on the HCL instrument than they spent responding to items on the LCL instrument. Although there were relatively few students who did not respond to any of the items on the instruments, the proportion of students who did not respond to any of the items on the HCL instrument was significantly higher than the proportion of students who did not respond to any of the items on the LCL instrument. Additionally, the proportion of items that students skipped on the HCL instrument was significantly higher than the proportion of items that students skipped on the LCL instrument. Finally, while the total number of students who displayed response patterns indicative of the use of response sets was similar for those answering items on both instruments, the types of aberrant response patterns that those two student groups exhibited differed.

This research builds on the work of researchers who have conducted studies examining the role of cognition in survey responding and educational psychologists who have proposed theories of cognitive load and carried out studies to measure it. My studies serve to bridge these two disparate bodies of research, drawing upon key findings from both literatures. The theoretical model I proposed represents a first attempt to identify survey design guidelines and recommendations related to the four steps in Tourangeau's model of the cognitive processes involved in survey responding. The model could prove very useful to practitioners if, over time, researchers can provide evidence that employing this model leads to reduced cognitive load and

improved survey outcomes. While I was only able to test some of the guidelines and recommendations in the model, the initial results seem promising, but researchers will need to conduct additional studies to test the efficacy of the full model when applied in varied contexts.

I. INTRODUCTION

A. Background and Statement of the Problem

The use of surveys is prolific in our society. From formal survey research to answering a few questions after a conversation with customer service, we are all inundated with requests to complete survey instruments. While these surveys come from countless sources, I am particularly interested in surveys that are used for the purposes of assessment in higher education. Because experts in survey research typically do not design these surveys, the quality of the surveys higher education professionals use varies widely. As the demand for higher education to prove its value increases, it is particularly important that higher education professionals attend to the quality of the surveys they use to assess student learning outcomes.

Accountability is becoming an increasingly important issue in American higher education. Several highly publicized events over the past decade have pushed the issue into the public spotlight. The reauthorizations of the Higher Education Opportunities Act in 2004, 2008, and the pending reauthorization, has sparked debate among politicians, university officials and the public about the meaning of accountability and what the American public has the right to expect from higher education. The graduation rates for higher education highlight the need for accountability. According to the National Center for Education Statistics, only 59% of students who begin an undergraduate degree graduate in the first six years, 57% at public institutions and 66% at private institutions (U.S. Department of Education, 2014). The push to change the face of accountability in American higher education began in the mid-1980s (Astin, 1990; Ewell, 1991) and has continued to grow and evolve through the reauthorization of the Higher Education Opportunities Act, the development of several commissions that focus on the state of higher education in the United States, and the publication of *Committing to Quality: Guidelines for*

Assessment and Accountability in Higher Education (New Leadership Alliance for Student Learning and Accountability, 2012). Ongoing need to reauthorize the Higher Education Opportunities Act continues to keep higher education accountability in the public spotlight.

For the 2008 reauthorization of the Higher Education Opportunities Act, former Secretary of Education, Margaret Spellings, created an advisory panel, the Commission on the Future of Higher Education, which gave its recommendations for accountability in higher education in the fall of 2006. Although this panel did not, as many people in the field of higher education feared, recommend mandatory standardized testing, its recommendations did spark debate about the issue of accountability in higher education (Fischer, 2006). In 2008, the government passed the Reauthorization of the Higher Education Opportunities Act (U.S. Department of Education, 2008) which affirmed individual institutional control of its instructional content and curricula.

In 2012, the New Leadership Alliance for Student Learning and Accountability published four guidelines for assessment and accountability in higher education: (a) set ambitious goals, (b) gather evidence of student learning, (c) use evidence to improve student learning, and (d) report evidence and results. Over 25 different organizations, including the Council of Regional Accreditation Commissions, endorsed the New Leadership Alliance's guidelines. This document set a clear expectation for assessment and provided guidelines for assessing student learning outcomes in higher education.

In July of 2013, all higher education regional accreditation bodies endorsed the Principles for Effective Assessment of Student Achievement (*Principles for Effective Assessment of Student Achievement*, 2013). These principles stress the need for institutions to define and assess educational outcomes for each of their programs.

The need to reauthorize the Higher Education Opportunities Act again in 2015 has sparked continued debate among politicians, the public, and higher education professionals about accountability. It is likely higher education accountability will be near the top of the agenda for this round of reauthorization (Stratford, 2014; "What You Need to Know about Reauthorization," 2013). In February of 2013, President Obama proposed that either changes to the current accreditation system be made or that a new system for distributing federal financial aid relying on "performance and results" be instituted. The college scorecard he proposed during his 2013 State of the Union Address for public accountability focused on cost, time to graduation, and post-graduation employment (White House, 2013).

With the huge number of students pursuing some type of higher education in the United States, accountability will remain an important issue for politicians and the public. To examine this issue fully, it is important to look at the sources of the increased demand for accountability in higher education and the history of accountability in higher education, including current practices.

The demand for increased accountability in higher education comes from two distinct sources: those external to the university, and internal sources. Historically, higher education has not had to demonstrate its effectiveness to its various stakeholders (Ewell, 1991). However, pressures for increased accountability now come from a variety of stakeholders. According to Kearns (1998), these stakeholder groups include (a) students and parents who shoulder an increasing financial burden for higher education, (b) politicians who are looking for a return on the money they invest in universities, (c) donors who want to see proof of the institutions' educational effectiveness, (d) alumni who worry about the economic value of their degrees, (e) special interest groups, and (f) the public. Because stakeholders support higher education, it is

important that colleges and universities respond to their demands for accountability for institutions' educational products (Burke & Modarresi, 2000).

As the business of higher education has matured, the state governments' demand for efficiency and effectiveness has increased as well (Van Valey, 2001). As higher education shifted from education solely for the country's elite to education for the masses (Neave, 1980; Van Valey, 2001), the number and diversity of students in colleges and universities increased greatly. This shift is due to a number of factors including the GI Bill and a public expectation that education is a democratic right and should be available for all (Ewell, 1991). Because the budgets for higher education have not increased at the rate of student enrollments, this sudden increase in student population means that universities must now do more with fewer resources. State governments choosing to invest more money in other public programs and services, like health care, welfare, and primary and secondary education, leaves little room to expand budgets for higher education (Ewell, 1991). Most colleges and universities now have larger class sizes, aging facilities, and decreased individual attention to students due to limited growth in faculty sizes (Ewell, 1991; King Alexander, 2000; Neave, 1980). This reduction in resources causes major stakeholders, such as the public and representatives in the state and national governments, to fear that educational quality and academic performance are also decreasing. As higher education expands, it is the institutions' responsibility to provide the public with evidence that they are meeting the same rigorous academic standards as in the past (Hodson & Thomas, 2003). State governments are very concerned about this perceived reduction in quality and are now pushing the issue of accountability in higher education to the front of their political agendas. For example, Scott Walker (R–Wisconsin) has proposed tying state funding for higher education to performance (Kingkade, 2012) and, borrowing language from Jeb Bush (R–Florida), argued that

Wisconsin "shouldn't be paying for butts in seats; we should be paying for outcomes" (Kingkade, 2012, para. 4).

The general public also places a great deal of pressure on institutions to demonstrate their educational effectiveness. This is due to several factors. The first is that higher education has not historically been open with the public about what it does and how it does it. As King Alexander (2000) explained, there is a divide between higher education and the external community. Traditionally, the public perceived higher education as secretive and not responsive to societal demands (King Alexander, 2000). This perception of aloofness on the part of higher education severely damaged the public trust in higher education, especially at a time when higher education was claiming more complex benefits from a university education (Ewell, 1991). Mistrust in the system increased the public's need for proof of educational effectiveness from institutions.

It is important that higher education address this issue by not only increasing its accountability to the public, but also by working to earn the public's trust. The public needs to see that higher education is responsive to societal demands and needs (Aper & Hinkle, 1991; Kearns, 1998; Leveille, 2005; Neave, 1980). As Leveille (2005) put it, "public trust and confidence in higher education are among its most important assets" (p. 11). Higher education serves the public and has to hold itself accountable to them by demonstrating that universities and colleges produce well-educated citizens.

The public also mistrusts the old "self-review" processes institutions have historically used to demonstrate accountability (Aper & Hinkle, 1991; Elton, 1988; Van Valey, 2001), resulting in an increased demand for accountability methods that allow students to compare one institution to another. This coincides with the public's perception of themselves as consumers.

Higher education is a public good. Students and parents feel like they are customers; and as customers, they want not only to have options in higher education, but also to have the opportunity to make informed decisions when selecting among those options. It is important that higher education provide students and parents with "consumer information" (Aper & Hinkle, 1991). The public wants universities to ensure that their students are receiving a good value for the money they invest.

In reaction to increasing demands for accountability, universities and colleges employ many different strategies to assess themselves. These activities were historically fairly similar; but as the demand for accountability increased, so did variety and creativity in assessment methods. A primary means for conducting assessment activities is survey methodology. According to a recent study, surveys are the most common assessment methodology utilized in higher education (Kuh, Janikowski, Ikenberry, & Kinzie, 2014). Administrators of academic programs frequently survey their current students, recent graduates, and those who employ their graduates to gather evidence that students have attained desired learning outcomes. Commercially available survey instruments are frequently very expensive and do not gather the specific information programs seek to provide evidence of student learning. For this reason, academic programs often write their own survey instruments. However, faculty and academic staff who want to write their own surveys may be overwhelmed by the amount of literature that exists on the topic of survey design and unsure of what or how much to read, given the limited amount of time they have to devote to this research. Those who do delve into the literature will find seemingly conflicting recommendations for best practices in survey design.

To illustrate, we will look at the issue of optimal number of response options to include in a rating scale. When looking for advice on the number of response options to include, faculty

will find that increasing the number of scale points can increase reliability (Jahoda, Deutsch, & Cook, 1951; Lehmann & Hulbert, 1972). However, there is a limit to how much one can increase reliability by including additional scale points. Some have argued that a scale having only two points (Alwin & Krosnick, 1991; Andrews, 1984) can produce reliable results, while others have argued that the optimal number of points for a rating scale is three points (Smith & Peterson, 1985), five points (Jenkins & Taber, 1977; Lissitz & Green, 1975; Wejters, Cabooter, & Schillewaert, 2010), six points (Kamorita & Graham, 1965), seven points (Andrews & Withey, 1976; Miller, 1956), or up to nine points (Alwin & Krosnick, 1991). To add to this confusion, there are arguments against using some of these numbers of scale points as well. For example, Andrews (1984) found that 3-point scales were actually less reliable than 2-point scales, but more reliable than 4-, or 5-point scales; and Smith and Peterson (1996) found that 7-point scales contain more error than 3-point scales. However, Andrews and Withey (1976) found that 7-point scales were better able to measure differences in respondents' feelings than 3-point scales. Someone referencing the literature to determine the number of scale points to include in a survey instrument may wonder how the literature can offer such different guidance. The truth is that each of these recommendations may be valid for certain conditions (based on respondent attributes, nature of the items in the survey, length of the survey, etc.).

However, navigating the literature to determine the most appropriate number of scale points to include in a survey instrument depending on the specific contextual elements of an individual survey is very difficult for those practitioners who attempt to base their designs on the literature. This is an especially salient point considering practitioners do not generally have the time or resources necessary to devote to a thorough and thoughtful examination of the full body of literature on the topic of number of scale points to include in their survey instruments. This is

only one consideration in designing a survey instrument; this type of confusion exists for essentially every aspect of survey design, including the wording of item stems, the design of scaled response options, the order of items presented on a survey, and others.

Given the challenge to practitioners of using the full body of literature to inform good survey design, what theoretical framework can best inform the many decisions a designer needs to make to create a survey instrument? One potential theoretical framework is based on the concept of *cognitive load*, or the amount of cognitive effort required for someone to respond to a survey instrument.

B. Purposes of the Studies

The purposes of the studies I conducted were to (1) create a succinct and easy-to-understand model that practitioners can use to design survey instruments, and then (2) test the efficacy and utility of a portion of that model. I based my model on a theory for reducing cognitive load in survey instruments (and in items on those survey instruments) and on findings from the survey design research literature. I then used the model to design online survey instruments to measure the level of agreement that students indicated when they responded to a series of statements regarding satisfaction with their college experience.

For my first study, I created a survey instrument that presented students with two versions of each item. The two versions measured the same content but differed in their cognitive load. (My theoretical model made clear the design principles that I followed to create the two instruments.) One was a high cognitive load (HCL) version, and the other was a low cognitive load (LCL) version. Students read the two versions of each item and then decided which version required more mental effort to respond.

In a second study, I randomly assigned students to complete a survey that contained all the HCL versions of the items or a survey that contained all the LCL versions of the items. If the survey design model were working as intended, then administration of the two instruments should have resulted in differential outcomes. I measured efficacy and utility in terms of several survey outcomes.

The first outcome was the *amount of mental effort* required to respond to high cognitive load (HCL) versions of items and low cognitive load (LCL) versions of those same items. If the survey design model were working as intended, then students should have reported that it required more cognitive effort to provide a response to the HCL version of each item than to the LCL version. That is, I hypothesized that, for each item, the proportion of students who selected the HCL version of the item as more cognitively demanding than the LCL version would be statistically significantly different from the proportion expected if students were to respond randomly when making their selections (i.e., 0.50).

The second outcome was the *amount of time* that students spent responding to the HCL instrument and to the LCL instrument. If the survey design model were working as intended, then there should have been differences in the amounts of time needed to complete each instrument. That is, I hypothesized that, on average, it would take longer for students to complete the HCL instrument than to complete the LCL instrument.

The third outcome was the *survey response rate* (i.e., the number of eligible college students who completed an instrument). If the survey design model were working as intended, then the number of completed instruments should differ. That is, I hypothesized that more students would complete the instrument that was built from the model.

The fourth outcome was *item response rate* (i.e., the number of eligible college students who provided a response to an item on an instrument). My hypothesis was that students would complete more of the items that appeared on the instrument built from the model.

Researchers who carry out studies on surveys frequently identify survey and item response rates as important indicators to monitor. Reducing the cognitive load of survey items and instruments may result in higher response rates. While recent studies indicate that the representativeness of the sample responding to a survey is more critical than the overall response rate (Krosnick, 1999; Visser, Krosnick, Marquette, & Curtin, 1996), Cook, Heath and Russel (2000) emphasized the importance of increasing a survey's response rate because it is often difficult to determine the extent to which the sample of respondents is actually representative of the population for whom the survey was intended.

Low response rate may result in several types of survey error. One potential threat is *coverage error*, or error that occurs because those who did not respond to a survey (or to an item on a survey) may differ in significant ways from those who did respond to the survey (or to an item on a survey) (Dillman, Smyth, & Christian, 2009). A related risk is *non-response error*, which occurs when the group who respond to a survey instrument (or item) is different from the group who did not respond to the survey instrument (or item) in ways relevant to what the instrument (or item) is measuring (Dillman et al., 2009).

The fifth outcome that I proposed to focus on was *response bias* (i.e., a systematic tendency to respond to items on a survey on some basis other than the content of the items). A primary form of response bias is *response sets*. As the term indicates, response sets are particular sets, or patterns, of responses a respondent uses when completing a survey instrument, regardless of the content of the survey.

If the survey design model were working as intended, then the number of students exhibiting response sets would differ across the two instruments. That is, I hypothesized that fewer students should show evidence of using response sets when completing the instrument built from the model.

Researchers who carry out studies on surveys frequently identify respondents' use of response sets as an important indicator to monitor. Reducing the cognitive load of survey items and instruments may decrease the use of response sets. When the task of responding to a survey requires more cognitive effort than survey respondents are willing to exert, respondents may be more likely to use response sets as a shortcut (Krosnick, 1991). Couch and Kenniston (1960) warned that the use of response sets may result in highly positive inter-correlations between ratings for all items on a survey, even when pairs of items are reversed (i.e., a pair of items that ask about the same concept, but one item is couched in positive wording while the other item is couched in negative wording).

The simplest type of response set in survey instruments using scales is choosing the same response option for every survey item. Another response set is *central tendency*, or the tendency to choose neutral options for every item and not use the full response scale. Krosnick and his colleagues (Krosnick et al., 2002) suggested this is a way in which respondents react to survey items and instruments that are too cognitively demanding. Another common type of response set is the tendency for respondents to choose an "agree" option when responding to all survey items, regardless of their content (Couch & Keniston, 1960; Greenleaf, 1992a; Wejters et al., 2010). The most extreme expression of an agreement response set may be detected when respondents choose an "agree" option for reversed items (Wejters et al., 2010). The tendency to agree may result from having a personality trait of wanting to be an agreeable person (Couch & Kenniston,

1960). Greenleaf (1992a) found evidence that a tendency to agree may not cause bias, but cautioned that in short surveys[1], this finding may not hold. Surveys that include balanced response scales, with an equal number of positive and negative response options, may reduce the effect of response sets (Couch & Kenniston, 1960).

An additional type of response bias, *extreme response styles* (ERS), is the tendency for a respondent to use the extreme ends of a scale, rather than responding with more moderate options, regardless of item content. The use of ERS may be related to individual respondent characteristics, such as age, education, and income (Greenleaf, 1992b). Greenleaf found that increases in age and decreases in education level and income led to increased use of ERS. The use of ERS in a survey causes bias by artificially increasing the standard deviation, or dispersion of item responses.

C. Definitions of Key Terms

I built a theoretical model for reducing cognitive load in survey items and instruments. For this study, I defined *cognitive load* as the amount of cognitive effort necessary to respond to survey items and instruments. The maximum cognitive load for an individual is limited by that respondent's working memory (Paas, Renkl, & Sweller, 2003). More specifically, the cognitive load imposed by a survey item cannot exceed the limits of the respondent's working memory (Paas et al., 2003) without the respondent either taking a shortcut or skipping the item altogether (Krosnick, 1991).

In this dissertation, I will use the word *survey* as a verb to signify requesting information, such as opinions, attitudes, or beliefs from a group or groups of interest. A *survey instrument* is the tool one uses to collect the desired information; the creation of this tool is *survey design*.

[1]Greenleaf did not define what he meant by a "short" survey; however, the survey in his study, which was not "short," contained 224 items

When building the survey design model, I focused on several *elements of survey construction*, including *visual design elements* and the *order of presentation of items* in a survey instrument.

In my studies, I considered how cognitive load might affect several different *elements of a survey instrument*. The first element was an *item*, or one request for information on a survey instrument, containing both a stem and response options. The *stem* is the prompt for information, and the *response options* are the possible ways in which a respondent can provide the information.

Each person who provides information on a survey instrument is a *respondent*. A *survey designer* is a person or group who creates a survey instrument, and a *survey administrator* is a person or group who uses a survey instrument to collect information from a particular population or sample.

D. Significance of the Research

Currently, there is no theoretical model that practitioners can use to design survey instruments. If practitioners are not already familiar with the wide-ranging survey design literature, they will have to spend a lot of time and effort reading the literature to appreciate the subtle nuances involved in implementing the myriad recommendations and best practices in survey design. Given that many practitioners are unlikely to have either the time or the motivation to read and study survey design to the extent necessary to master the science and art of survey design, they may find it useful to have a practical, easy-to-understand model to follow when creating their survey instruments, especially if the use of this model for reducing cognitive load could be shown to lead to improved survey outcomes.

E. Research Questions

The overarching question that focused the research was this: Can using a theoretical model for reducing cognitive load to create a survey instrument improve survey outcomes? The more specific questions that I proposed tested a portion of the theoretical model in terms of its ability to improve survey outcomes. These research questions addressed student performance on the survey instrument as a whole (i.e., on a theoretically low cognitive load (LCL) instrument and on a theoretically high cognitive load (HCL) instrument).

1) Will students report that it takes more mental effort to respond to high cognitive load (HCL) items than to low cognitive load (LCL) items?
2) Will students spend more time responding to a HCL instrument than to a LCL instrument?
3) Will more students complete the LCL instrument than the HCL instrument?
4) Will students respond to more items on the LCL instrument than to items on the HCL instrument?
5) Will more students use response sets when answering items on the HCL instrument than when answering items on the LCL instrument?

II. REVIEW OF LITERATURE

A. Organization of the Literature Review

This literature review begins with a look at the early research on the role of cognition in survey design. Beginning in the 1970s and 1980s, many researchers began moving away from relying upon stimulus and response models to explain behavior. Instead, they began to explore the use of sophisticated models of cognition to explain behavior, models that focus on individual agency in response to various stimuli. This opened the door for researchers interested in survey design to investigate the role of individual cognitive processes in responding to surveys.

Next, I explore the cognitive load literature, including the literature on the theory of cognitive load and its measurement. This literature helped inform my decisions regarding how best to measure cognitive effort and its effects on survey respondents. I then look at the literature on modeling the cognitive processes involved in responding to surveys, seeking to determine how researchers have defined and measured optimal cognitive effort.

To help me develop a theoretical model to reduce cognitive load in surveys, I continue with a look at the cognitive effects on survey responding of key elements of survey design and construction. I compare two different, but related, theoretical models that seek to identify and describe the cognitive processes in which respondents engage as they respond to a survey. The chapter concludes with my proposal for a theoretical model to reduce cognitive load in surveys.

B. Early Cognitive Research in Survey Design

Survey researchers started exploring the influence of cognition in surveys in the 1970s and 1980s (Jobe & Mingay, 1991). In the 1970s, researchers began to look at the effects of survey task burden on survey outcomes (Cannell, Miller, & Oksenberg, 1981; Oksenberg & Cannell, 1977; Schuman & Presser, 1977; Sudman & Bradburn, 1974). In 1974, Sudman and

Bradburn compared task characteristics and respondent characteristics and found that elements of the task that they asked respondents to complete had more impact on the accuracy of their survey responses than those respondents' individual traits.

In 1987, Edward Blair and Scot Burton carried out a seminal study, exploring three different hypotheses that support the argument that increasing the cognitive load of behavioral frequency items decreases response accuracy. (*Behavioral frequency items* are items that ask respondents to report how often certain events occurred.) The researchers reported three related findings. First, the researchers found that the larger the number of events that the respondents reported, the less likely they were to recall the events from memory and count them, which they considered the most accurate way of reporting the frequency of the events. In these cases, the respondents were more likely to use other, less precise methods, such as estimating. The researchers also found that the longer the timeframe of focus for an item, the less likely respondents were to actually count the number of events they reported. Finally, Blair and Burton looked at the effect of the more ambiguous wording "how often," in comparison to the more specific wording "how many times." The results from their analysis supported their hypothesis that using the more specific wording, which would theoretically require less cognitive effort to interpret, led to more accurate reports of behavioral frequency.

C. Theory and Measurement of Cognitive Load

Theoretical models of cognitive load are based on concepts from cognitive psychology, later developed into cognitive load theory by researchers in educational psychology. Paas and Van Merrienboer (1994a) developed a model for cognitive load and its measurement grounded in a cognitive load theory that Sweller (1988) proposed. According to their model, cognitive load has both *extrinsic* and *intrinsic elements* (Paas, Tuovinen, Tabbers & Van Gerven, 2003; Paas &

Van Merrienboer, 1994a). *Extrinsic elements*, collectively termed the *mental load*, are those the task imposes, and *intrinsic elements* are dependent on the individual characteristics of the person completing the task (Paas, 1992; Paas et al., 2003; Paas & Van Merrienboer, 1994a). These researchers also describe a third element of cognitive load: *extraneous*, *or unnecessary cognitive load.* This type of cognitive load collectively describes all elements of cognitive load that are not necessary to complete the task. Examples include spelling errors in an item stem, unnecessary use of unfamiliar vocabulary, and unnecessarily complex sentence structures. These three sets of elements combine to create the *overall cognitive load of the task* (Paas et al., 2003; Paas & Van Merrienboer, 1994a). Paas and Van Merrienboer (1994a) further broke down the concept of cognitive load into (a) those elements that are *causal*, including task characteristics, respondent characteristics, and the interaction between the two, which they defined as the *interaction element*, and (b) those that are *measurable*, which they referred to as the *assessment dimension*. These researchers pointed out that survey designers can only directly control task characteristics, but they have no control over respondent characteristics. The implication for survey designers is that their greatest potential for reducing cognitive load is in focusing on task characteristics of survey items and instruments. Examples of task characteristics include format, complexity, use of multimedia, and time required to complete.

One may reasonably ask, why measure cognitive load as it relates to survey response? Researchers theorize that if a survey item is too cognitively taxing, respondents will expend progressively less cognitive effort to produce a response (Cannell et al., 1981), resulting in a decreased amount of information the survey will have the ability to collect. If this is true, surveys requiring a high degree of cognition to process and respond to would likely result in less thoughtful responses and a reduction in the amount of information available from those survey

instrument's results. Theoretically, reducing the cognitive load of a survey should result in an instrument capable of collecting more and better information.

Cognitive load consists of three assessment dimensions: (a) *mental load*, imposed by the environment and not directly controllable by survey designers; (b) *mental effort*, or the cognitive capacity and mental resources devoted to the survey response task; and (c) *performance*, or the final products of the task (Paas & Van Merrienboer, 1994b). Measurement of cognitive load may involve assessing one, or some combination, of these aspects. However, some researchers argue that the "essence" of cognitive load is mental effort; therefore, it may be the best index for the measurement of cognitive load (Paas & Van Merrienboer, 1994a, 1994b).

Measurement of cognitive load may be either analytical or empirical. Analytical methods to estimate mental load may employ various techniques, such as task analysis or statistical modeling of response patterns, followed by the collection of additional subjective evidence using expert opinions of the mental load items should impose (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Empirical models attempt to estimate cognitive load more directly. Paas and Van Merrienboer (1994a) described three potential methods for empirically measuring cognitive load: subjective data, performance data, and physiological data. Some researchers measure mental load using subjective measures, such as asking respondents to report the amount of mental effort exerted when responding to a survey item (Paas & Van Merrienboer, 1994a). While Paas and Van Merrienboer argued that sufficient evidence exists to assert that respondents can reliably access and report the amount of cognitive effort they exerted, Nisbett and Wilson (1977) argued that respondents may not have access to the cognitive processes they use in responding to survey items. Instead, respondents may only have access to their preconceived notions of the processes they feel they should be using to respond to that survey item. That is,

respondents may have access to the cognitive processes they used to respond to a survey item or items, but only in the event that there is no disconnect between the cognitive processes they think they should be using to respond and the cognitive processes they actually used to respond. These researchers' differences of opinion regarding whether respondents can appropriately access their cognitive processes may reflect differences in the aspects of cognitive load that the researchers measured: Nisbett and Wilson (1977) focused on mental load (i.e., the cognitive processes the respondent is employing), while Paas and Van Merrienboer (1994b) focused on mental effort (i.e., the amount of effort the respondent has to put forth to respond to a survey item or items). These findings would seem to suggest that cognitive effort is more reliably accessible to respondents than the specific cognitive processes they employed.

Other researchers measure mental load using physiological measures, such as pupil dilation and heart rate monitoring (Paas & Van Merrienboer, 1994a). However, these methods are necessarily very physically intrusive and may not reliably measure subtle difference in cognitive load (Paas, Tuovinen et al., 2003).

Yet another method involves employing empirical methods that combine the gathering of the respondents' subjective ratings of their mental effort along with data on their performance of the task and physical data such as heart rate or pupil dilation to estimate cognitive load (Paas, Tuovinen et al., 2003).

D. Modeling the Cognitive Processes Involved in Survey Responding

Theorists have proposed several different models to characterize the cognition involved in responding to a survey instrument. I will focus on two related models. The first model is the seven-step, question-answering process that Cannell and his colleagues (1981) outlined. I will look at this model because most of the subsequent models researchers have proposed are based

on this model. The second model is the most frequently cited and utilized cognitive model in survey literature, Tourangeau's 1984 model of cognitive processes in survey response.

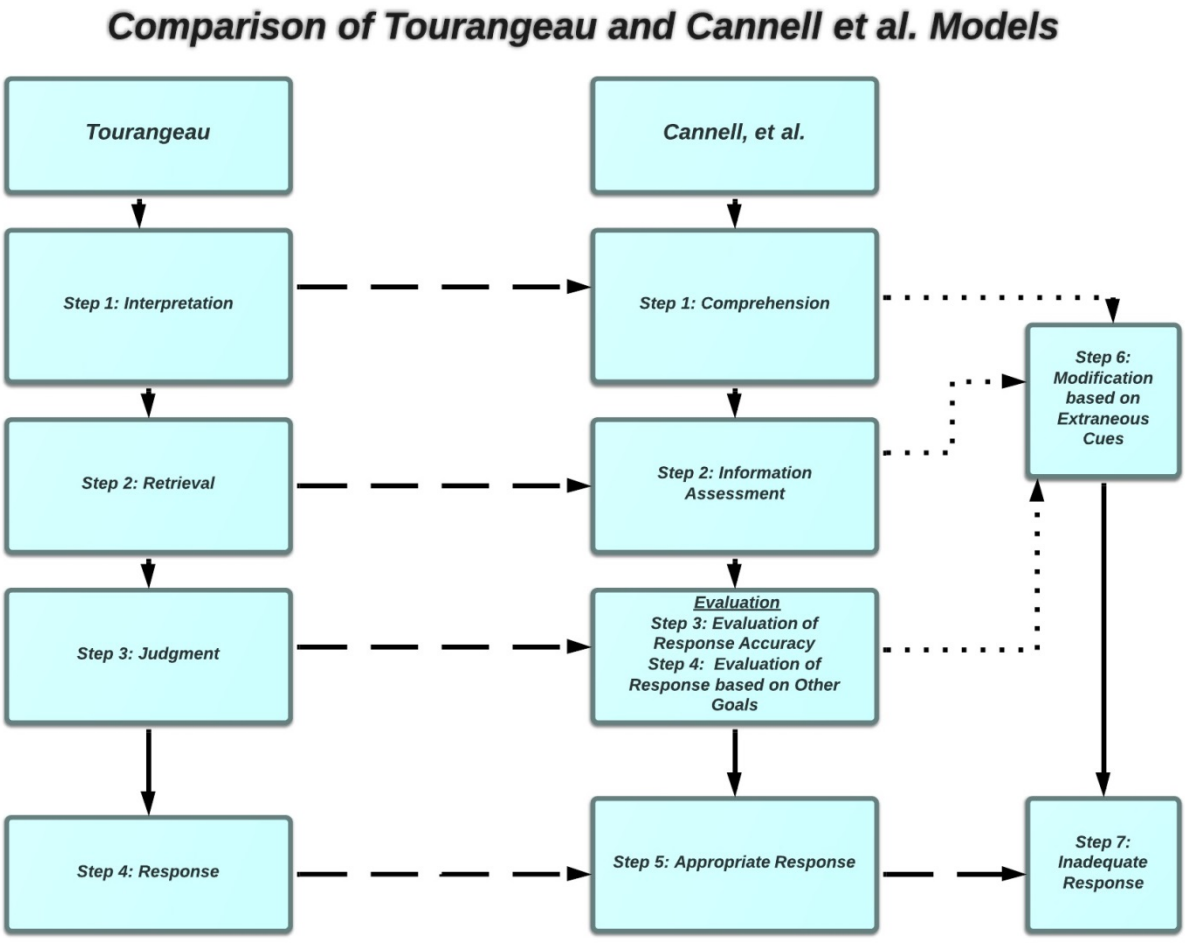

Figure 1. Comparison of the models of Tourangeau and Cannell, Miller and Oksenberg.

1. Cannell, Miller, and Oksenberg's seven-step, question-answering process

Charles Cannell, Peter Miller, and Lois Oksenberg (1981) presented a model that describes the cognitive processing in which respondents engage when they are responding to a survey item. These researchers proposed an optimal five-step process for cognitively processing a survey, with two additional steps that can distract respondents from the survey response task

and cause unnecessary or incorrect cognitive processing, as well as inaccurate or incomplete responses.

Step 1 (see Figure 1) involves comprehension of the question. Comprehension requires the respondent to understand each of the words used in the item stem, the concept that the stem references, and the sentence structure that the survey designer used. This step also includes respondents' interpretation of the question. There may be multiple possible interpretations of the question, in which case the respondent needs to consider carefully each of the possible interpretations to determine which one is most appropriate.

Step 2 requires respondents to go through a process of retrieving and evaluating the information necessary to respond to the question. Respondents must retrieve information from their memories, consider the relevant context for that information (i.e., determine whether or not the information is relevant given the context of the question), and organize the information to arrive at an appropriate response.

Step 3 in the process is the evaluation of the response to determine if it is adequate, given the question. If the respondent determines the response is not adequate, he or she may need to go back to perform one or more components of Step 2 again. A second type of evaluation occurs in Step 4. The respondent engages in a psychological evaluation to determine whether the response may pose a psychological threat, such as to the respondent's self-perception, or to how the respondent desires others to perceive him or her. If the respondent determines the response is sufficiently non-threatening, he or she then gives the most accurate response possible (Step 5).

While these five steps represent the ideal response process, respondents may, for various reasons, become distracted from this path at any step along the way. The respondent may deviate to Step 6 in this process and alter his or her response based on cues external to the

process. For example, the respondent may take cues from an interviewer, become distracted by other questions in the survey instrument, or view an accurate response as potentially embarrassing. Once a respondent deviates to Step 6, he or she will produce an invalid or biased response (Step 7) instead of giving a truthful, well-considered response (Step 5).

2. Tourangeau's four-step process

In a 1984 review of literature, Roger Tourangeau proposed a similar, but simplified model for characterizing cognition's role in survey responding. This model specified four distinct steps in cognitive processing necessary to respond to survey instruments.

According to this model, Step 1 in cognitive processing is interpreting the item stem. A respondent must understand the meaning of the question. Next, the respondent must retrieve from memory the necessary *information* (e.g., attitudes, interests, beliefs, opinions) to formulate a response. In Step 3 in this process, the respondent must make a judgment, based on the information that he or she has accessed before providing a response in Step 4.

Each of the steps in this process requires distinct cognitive processes. In Step 1 (interpretation), the respondent must determine what information the question is seeking in his or her response. Ambiguous wording or unfamiliar vocabulary can increase the amount of cognitive effort necessary for a respondent to interpret the question (Cannell et al., 1981; Dillman et al., 2009; Fowler, 1995). Cognitive load also depends on familiarity with the item content. When a respondent is unfamiliar with the issue presented in the item stem, he or she will find it more difficult to determine an appropriate frame of reference to understand the question (Tourangeau, 1984).

In Step 2 (retrieval), the respondent will have an easier time accessing from memory the relevant information for issues with which he or she has more familiarity. Respondents are

unlikely to retrieve every piece of information they have about a particular issue, so they will likely be selective in what they retrieve, which will over-represent the information that is most readily accessible (Tourangeau, 1984).

In Step 3 (judgment), cognitive demand will depend on the saliency of a respondent's pre-existing information about the issue of interest. For example, if the respondent already has a well-developed opinion on an issue, it may require very little cognition for the respondent to select a response option. In fact, the respondent may be able to simply retrieve a pre-existing opinion from long-term memory, putting no demands on working memory. However, survey designers may inappropriately assume this will always be the case (Tourangeau, 1984).

Survey items frequently ask for opinions or attitudes about which the respondent has not previously made a judgment. In this case, the respondent will likely have to retrieve from memory the most closely related information, and then engage in a process of integration to arrive at his or her final judgment, requiring a higher degree of cognitive processing (Tourangeau, 1984).

In Step 4 (response), the respondent needs to take the final judgment and map it to the available response options. To do this, the respondent will need to interpret the response options and then decide which one is most appropriate, given his or her final judgment. After making this decision, the respondent may also check the response for consistency with his or her previous responses to other survey items, cues from the survey instrument, social influence or other contextual issues, which may differ based on the culture of the respondent (Johnson, O'Rourke, Chavez, Sudman, Warnecke, & Lacey, 1997).

E. Optimal Cognitive Effort in Survey Design

To evaluate the role of cognition in survey design, researchers have defined optimal cognitive effort in responding to survey instruments (e.g., Cannell et al., 1981; Krosnick, 1991). Optimal effort involves two different considerations: (a) what is the optimal cognitive effort, and (b) how does one define deviations from this ideal?

1. Optimization

The optimal exertion of cognitive effort by survey respondents, or *optimization* (Krosnick, 1991), occurs when respondents exert the cognitive effort necessary to fully engage in each of the five steps in the question-answering process that Cannell and his colleagues (1981) proposed, or in each of the four steps in Tourangeau's (1984) model. However, if the cognitive demands of a survey are too great, respondents may take shortcuts in one or more of the cognitive processes (Cannell et al., 1981), or *satisfice* (Krosnick, 1991).

2. Satisficing

Satisficing, a combination of the terms *satisfy* and *suffice*, refers to the minimum amount of effort one needs to exert to fulfill a need (Simon, 1956). In the case of survey response, it is the minimum amount of effort a respondent needs to put forth to respond to a survey item or instrument (Krosnick, 1991, 1999).

According to Cannell et al. (1981), the main threat to the validity of survey results is respondents who do not fully cognitively process the survey item, thereby distorting the survey data. The distortion may take two different forms: (a) the under-reporting of information, or *false negative;* or (b) the over-reporting of information, or *false positive*. A *false negative* occurs due to reduced cognitive processing in the second step of their model.

Krosnick's theory of satisficing is based on Tourangeau's (1984) model. Krosnick (1991) defined two different levels of satisficing: *weak* and *strong*. For example, a respondent who engages in *weak satisficing* may simply respond with the first option that seems rational instead of carefully going through each of the cognitive steps necessary to produce an optimal response (Krosnick, 1991). *Strong satisficing* occurs when a respondent fails to go through all of the steps, eliminating one or more to arrive at a response without exerting much cognitive effort. While Krosnick defined two levels of satisficing, he conceived of satisficing as forming a continuum, from weak to strong satisficing.

Krosnick (1991) theorized that respondents engaging in weak satisficing will show *primacy effects*, or the tendency to select the first options presented in a list, when completing a self-administered survey instrument. Respondents may give the most careful consideration to the first options and then progressively devote less cognitive effort as they continue to read options. Alternatively, respondents may not read the entire list of options and simply select the first that seems appropriate. Weak satisficers may also demonstrate *acquiescence response bias* (Krosnick, 1991), or the tendency to agree, because it requires less cognitive effort than disagreeing.

When using a common scale for a series of survey items, respondents may demonstrate strong satisficing by only considering the first item in the series of items and then using the same response for the remainder of the items that employ the same scale (Krosnick, 1991). This is an example of respondents using a *response set* (i.e., employing a common response regardless of the content of the individual items). Types of response sets include *yeasaying* (i.e., the tendency to agree regardless of content), *naysaying* (i.e., the tendency to disagree regardless of content), *standard deviation* (i.e., the tendency to either exhibit very little (or a lot) of variation in one's

selection of response options), and *extreme response style* (i.e., the tendency to only choose response options that appear at the ends of a scale). According to Krosnick (1991), another example of strong satisficing is choosing response options without considering either the item stem or the response options. Krosnick called this *mental coin-flipping*, or selecting responses randomly. One of the most pervasive forms of satisficing is *non-response*. Because skipping an item requires no cognitive effort on the part of the respondent, Krosnick considered this the strongest form of satisficing.

F. Cognition and Key Elements of Survey Design

To reduce the overall cognitive load of a survey instrument, survey designers need to consider the cognitive effects of every element of the instrument, including not only the *main survey components* but also *other significant survey design details* that the designer does not intend for respondents to pay attention to when they are responding to a survey instrument, but nonetheless are important. The *main survey components* include the items themselves (i.e., the *item stems* and the *response options)*, and the *survey instructions*. *Other significant survey design details* include both the *order of the items* and certain *visual design elements*.

1. Main survey components

a. Item stems

Item stems should make clear the information that respondents are to access from memory when they are formulating their responses (Dillman et al., 2009). The language used in the stem has a direct bearing on the cognitive effort needed to interpret it. Unfamiliar vocabulary, language, jargon, or acronyms require greater cognitive effort to read and properly interpret (Cannell et al., 1981; Dillman et al., 2009; Fowler, 1995).

Additionally, ambiguous words or phrasing (i.e., words or phrases that a respondent could reasonably interpret in multiple ways) increase the cognitive load of the item stem. Respondents need to determine which meaning the survey designer intended before they can select the most accurate response (Cannell et al., 1981; Dillman et al., 2009; Fowler, 1995). The process of interpreting the meaning adds to the cognitive effort necessary for not only understanding the stem but also selecting a response.

One example of ambiguity in an item stem is a survey item that focuses on multiple concepts, such as in a double-barreled item (Dillman et al., 2009; Fowler, 1995). This type of stem increases the cognitive effort required to respond to it because the respondent may need to access from memory different sources of information to respond to each concept; however, the respondent is forced to choose only one response to respond to both concepts, which may place the respondent in the uncomfortable position of having to reconcile potentially conflicting sources of information.

Additional sources of ambiguity in the item stem can increase the overall cognitive effort a respondent needs to expend to respond to the item. For example, unclear concepts in the stem (Cannell et al., 1981; Dillman et al., 2009) and complex sentence structures (Cannell et al., 1981; Dillman et al., 2009) increase the overall cognitive demand of a survey instrument. The respondent's cultural background plays a critical role as well, since an individual interprets concepts through the lens of his or her culture (Johnson et al., 1997).

Survey designers should avoid item stems that force the respondent to make any assumptions (Dillman et al., 2009; Fowler, 1995). For example, the stem "How much did this course help increase your understanding of African American history?" makes the assumption that the course increased the respondent's understanding. If the respondent does not feel that the

course increased his or her understanding of African American history, it will be cognitively taxing to choose an accurate response, and the respondent may choose to skip the item rather than select a response that is not correct.

A survey designer can also add cognitive load to an item stem by requesting information with which the respondent does not have primary experience. For example, stems based on hypothetical situations (Fowler, 1995) or second-hand experiences, such as something a respondent heard about rather than personally experienced (Fowler, 1995), are cognitively complex.

Additional cognitive effort is required to respond to item stems that contain double negatives. In this case, the respondent needs to disagree in order to agree with a statement (Dillman et al., 2009). For example, suppose that an item stem asks respondents "Do you believe ex-convicts should not be allowed to have gun licenses?" In order to agree with the concept that ex-convicts should be allowed gun licenses, the respondent needs to disagree with the original premise, creating a double negative. Making sense of a double negative requires an additional cognitive step in order to appropriately interpret and respond to the item. A survey designer can reduce the cognitive load of this stem by phrasing it "Do you believe ex-convicts should be allowed to have gun licenses?"

There is some disagreement in the literature regarding the effect of the length of an item stem on the cognitive effort respondents need to exert to interpret it. Some argue that a longer stem decreases the need for expending cognitive effort by better describing or reinforcing the information needed to respond to the item (Cannell et al., 1981). By contrast, others argue that shorter more succinct stems are easier for the respondent to read and interpret, decreasing the cognitive effort the respondent has to exert to respond (Dillman et al., 2009). A reasonable

compromise position might be that stems should be succinct but explicit. That is, when creating an item stem, the survey designer should aim for brevity but clearly indicate the information sought.

The type of cognitive activity in which the item's stem requests (or requires) the respondent to engage is also an important determinant of the cognitive effort needed to respond (Fowler, 1995). For example, the degree to which a stem requires a respondent to search his or her memory is an important factor in determining the amount of cognitive effort the respondent will need to expend to respond to the item (Cannell et al., 1981; Dillman et al., 2009; Fowler, 1995).

b. Response options

Responses to survey items are based on the content of the survey item and on an evaluation of the response scale itself (Couch & Kenniston, 1960). In their study of the psychological attributes of respondents, Couch and Kenniston (1960) found that respondents who tended to agree with survey items (as well as those who tended to disagree with survey items) used not only survey item content but also the response scale to arrive at a final response.

i. Number of response options

Much of the existing research supports the conclusion that increasing the number of response options increases the reliability, or consistency or responses, of an instrument (e.g., Bendig, 1953; Murphy & Likert, 1938; Peter, 1979). Offering too few response options may not provide respondents enough variability to map accurately their views onto the response scale (Alwin & Krosnick, 1991). However, there is also a practical need to balance the reliability gains one might obtain from adding response options against the increased burden that additional response options place on the respondents (Symonds, 1924). Increasing

the number of response options may reduce cognitive load by making it easier to select an option that most accurately fits the respondent's opinion. However, with each additional response option, the cognitive demand increases. The respondent must read and interpret the additional response options, and then process all the alternatives to select the most appropriate response (Krosnick, 1991). Additionally, increasing response options is only effective if each option is a distinct and meaningful division of the *concept of inquiry*, or topic for which a survey is requesting information (Alwin & Krosnick, 1991).

The number of response options may not be the most important factor in determining the reliability of a survey instrument. In a 1977 study, Jenkins and Taber studied different variables to determine their effects on reliability including the number of response options, the degree of covariance in the ratings that respondents assigned to items, and judgment precision. They found that while the number of response options had a significant impact on the reliability of the survey instrument, it had the least impact on the variables they studied. Cronbach also warned that increasing reliability does not necessarily increase validity, or accuracy of responses (Cox, 1980). Considering these findings, it seems reasonable to limit the number of response options to the smallest number required to cover the full range of the concept of inquiry to reduce the cognitive load, even at the potential expense of sacrificing some possible gains in reliability.

Cox (1980) broadly defined the optimal number of response options as the number of options necessary to offer the most relevant information in the set of options without so many that the information the respondent provides when responding to the item begins to degrade. In lay terms, this means "it depends." It depends not only on the content of the survey item (which places boundaries on the type and amount of information that can be collected), but also on the purpose of the scale (which places boundaries on the amount of systematic variation the survey

designer can reasonably expect to see in responses). The best surveys optimize systematic variance in responses and minimize random variance (Cox, 1980).

More specifically, Cox (1980) identified four different factors that a survey designer should consider when deciding on the number of response options to include in a survey instrument: (a) the ability of the respondents to select an appropriate response based on the stimulus (i.e., item stem) presented; (b) the capacity of the scale to collect information; (c) the amount of information that the stimulus (i.e., item stem) is able to elicit; as Cox stated "a respondent can transmit no more information than is found in the stimuli in the first place" (p. 411); and (d) the information that the survey administrator needs. Though it depends on the survey instrument, survey items, surveyed population, and information the survey administrator is attempting to obtain, the optimal number of response options may be four to five (Cox, 1980), depending on whether the survey includes a neutral response option.

ii. Inclusion of a neutral response option

Response scales that include a neutral response option assume that the respondent who chooses that option is truly neutral, having no opinion either positive or negative. When determining whether to include a middle, or neutral, response option in a rating scale, the survey designer must consider what it will mean when respondents select that option. Survey data analysts generally code a neutral response as signifying a number in the middle of a continuum from positive to negative. This becomes problematic when drawing conclusions based on the data analysis if a respondent's selection of the neutral category does not actually represent a true lack of an opinion.

Klopfer and Madden (1980) posited three different reasons for choosing a neutral response option on a rating scale: neutrality, ambivalence, and uncertainty. Respondents may

genuinely have no opinion due to lack of exposure to a relevant issue, or because they have never thought about the issue (Alwin & Krosnick, 1991). Alwin and Krosnick (1991) theorized that when a respondent is truly neutral with respect to an issue and the survey instrument offers no neutral response option, the respondent has two choices: (a) randomly choose one of the two response options that appear in the middle of the scale, or (b) skip that item.

Another possible reason for selecting a neutral response option is that respondents are ambivalent. A respondent may sometimes feel one way about an issue and at other sometimes feel another way, or a respondent may feel one way about one aspect of an issue and another way about a different aspect of that issue (Feick, 1989; Klopfer & Madden, 1980). Even a respondent who has a well-formed attitude may be ambivalent when overwhelmed by the cognitive task of determining which response option best reflects that attitude (Alwin & Krosnick, 1991).

Vague or indistinct response options may also introduce ambiguity. One example is a scale that includes both "infrequently" and "seldomly" as response options; it may be unclear to respondents which response option signifies less frequency. Unclear item stems can also introduce ambiguity if respondents are uncertain about the meaning of the item (Alwin & Krosnick, 1991).

Researchers studying the effects of including a neutral response option in survey items have reported some intriguing findings. Alwin and Krosnick (1991) theorized that survey instruments with no neutral response option would have lower reliability than those that include one. However, they found just the opposite; scales with no neutral response option actually produced higher reliabilities. Krosnick et al. (2002) reported that including a neutral response option did not increase the quality of the data as measured in either the consistency of attitudes over time or the statistical predictability of responses. These researchers argued that offering a

neutral response option encourages satisficing. In this case, a neutral response option is a handy way for respondents who are overly cognitively taxed to take a shortcut in their responses rather than a way to represent a truly neutral opinion (Krosnick et al., 2002). Krosnick et al. (2002) found that less educated respondents were more likely to use neutral response options than their more educated counterparts, supporting these researchers' hypothesis that respondents use neutral response categories as a cognitive shortcut.

Klopfer and Madden (1980) studied four possible reasons for why respondents may use the neutral response option: (a) *uncertainty,* or an inability to define their opinions; (b) *ambivalence*, or sometimes feeling one way about an issue while at other times feeling a different way; (c) *neutrality*, or not having an opinion one way or the other; and (d) *nonspecific*, or unable to decide. These researchers concluded that ambivalent respondents may be more likely to use the neutral response option than respondents who hold truly neutral opinions.

The decision of whether or not to include a neutral response option may depend on the type of bias that the survey designer hopes to minimize; inclusion of a neutral response option may increase a respondent's tendency to agree with all survey items but decrease the likelihood that a respondent will use extreme response sets (Wejters et al., 2010). While a neutral response option can be misused, Cox (1980) argued that it should be included when some respondents may reasonably be expected to have no opinion about an issue. Theoretically, inclusion of a neutral response option should reduce cognitive load when the survey designer reasonably expects that some respondents will not have an opinion about an issue, but may threaten the validity of the results from a survey by encouraging satisficing when the survey designer expects all respondents to have an opinion about the issue.

iii. Labeling response options

There are several different ways to label response options: (a) fully labeling the options by providing a descriptive label for each one, (b) providing descriptive labels for those options appearing at the upper and lower ends of a scale, (c) fully labeling the options by assigning a numeric value to each one, (d) assigning numeric values to those options appearing at the upper and lower ends of a scale, or (e) a using a combination of descriptive and numeric labels for the options. However, Alwin and Krosnick (1991) cautioned that providing numeric values alone may result in ambiguous response scales.

The effect of fully labeling scales is not completely clear. The use of fully labeled scales may result in higher reliability (Alwin & Krosnick, 1991) and lower extreme response bias (Wejters et al., 2010). However, the use of fully labeled scales may also increase acquiescence response bias (Wejters et al., 2010) and response time (Tourangeau, Couper, & Conrad, 2007). In a 2007 article, Tourangeau and his colleagues compared the response times for fully labeled response scales and response scales with only the response options at the extremes of the scale labeled. The amount of time required to respond to a survey item may be a good indication of the cognitive load of that item. The researchers found that fully labeled scales required more response time than scales in which only the response options at the extremes of the scale were labeled. However, they found that if all the items on the survey instrument used the same fully labeled scale, this effect faded as the respondent no longer required additional time to read and understand the response scale.

The addition of numeric values to response scales can change the meaning of the response options. Respondents interpret scales differently if numeric scales include both negative and positive numbers (as opposed to positive numbers only). Researchers reported that

respondents tended to assign slightly more positive ratings on average when they used scales with numeric labels containing both positive and negative numbers (e.g., -5 to 5) as opposed to scales with numeric labels containing positive numbers only (e.g., 0-10) (Schwarz, Knauper, Hippler, & Noelle-Neumann, & Clark,1991; Tourangeau et al., 2007) When numeric labels contained all positive numbers, respondents did not assign additional meaning to response options beyond the information they obtained from the descriptive labels for the options appearing at the extremes of the scale. However, when numeric labels contained both negative and positive numbers, responses indicated that the numeric labeling of the response options communicated additional information beyond that provided by the descriptive labels. Respondents tended to view the response options at the lower end of the scale as more negative, and therefore less attractive options, even though both scales contained the exact same descriptive labels (Schwarz, Knaupfer et al., 1991; Tourangeau et al., 2007). Schwarz, Knauper, and their colleagues hypothesized that this effect may be more pronounced in self-ratings where social desirability pressures motivate respondents to avoid especially negative ratings of themselves. Overall, these studies demonstrate that when engaged in the cognitive processes of interpreting and responding to a survey item, respondents use numeric labels as a source of information in addition to information that descriptive labels provide. This is especially true when descriptive labels are not clear (Schwarz, Knauper et al., 1991). Numeric labels may be helpful in labeling response scales, either alone, or in combination with descriptive labels. Because respondents use numeric labels as additional sources of information to interpret scales, these labels may also reduce cognitive load. However, if using numeric scales, it would be best to avoid the use of both positive and negative numbers, as this may create a positive response bias.

c. Instructions

Making clear to respondents a survey's purpose and the procedures they are to follow when completing the survey can help to reduce cognitive load. A lack of clear instructions increases the complexity of responding to a survey because the respondent may need to sort through alternative ways to progress through the survey and respond to its items, then decide on the one way the survey designer intended the respondent use (Paas & Van Merrienboer, 1994b). Cannell et al. (1981) found that clear survey instructions reduced respondents' cognitive effort by clarifying how the results from the survey would be used, and how they should approach the task of responding to the survey. This was especially true for highly educated respondents. For example, instructions may remind respondents to please consider each option carefully before responding to a survey item. This type of instruction may be especially helpful when a survey designer can foresee potential errors that respondents might be likely to commit when completing a survey, or incorrect interpretations that respondents might make when considering particular items on the survey instrument (Cannell et al., 1981).

G. Cognition and Survey Construction

In addition to the main components of a survey, certain elements of survey construction may also contribute to the overall cognitive load of a survey instrument. The orders of presentation for both items and response options contribute to cognitive load, as well as other visual elements of a survey instrument's design.

a. Item/response option order

i. Item order

The order of survey items may have different effects on responses. One possible outcome is *assimilation effects*, or the tendency to respond in a similar manner to a set

of items, paying little attention to each item's specific content (Wanke & Schwarz, 1997). By contrast, item order may also result in a respondent showing *contrast effects*, or the tendency to respond in a contradictory manner to two or more items that have very similar content (Wanke & Schwarz, 1997). Both of these outcomes affect the quality of the survey data because the responses are item content irrelevant.

Wanke and Schwarz (1997) argued that these effects stem from respondents' use of multiple sources of information as they seek to understand and respond to survey items. Ideally, respondents should gather all information for their interpretations and responses to survey items from the survey instructions, item stem, and their own memories. However, when respondents are interpreting survey items, they will use all sources of information available to them, even if the survey designer did not intend for all those sources to inform item interpretation (Couper, Tourangeau, & Kenyon, 2004; Wanke & Schwarz, 1997). Wanke and Schwarz pointed out that respondents generally want to be helpful and provide the information that the survey designer is seeking. Thus, they will tend to assume that every piece of information in the survey instrument is relevant, and that they should use all the information available to them as they formulate their responses.

When respondents are interpreting the semantic meaning of an item stem and the item is vague, ambiguous, or contains unfamiliar vocabulary, respondents may use the concepts that previous survey items recently primed to interpret the meaning of the current item, resulting in assimilation effects (Tourangeau & Rasinski, 1988; Wanke & Schwarz, 1997). Proximity can also cause *assimilation effects*; respondents will tend to assume items grouped together are related in content, and more distal items are less related (Wanke & Schwarz, 1997). However, if

the respondents become aware of assimilation effects, they may over-correct, resulting in *contrast effects* (Wanke & Schwarz, 1997).

Assimilation and contrast effects may affect respondents at the judgment stage when respondents are determining the meaning of an item in order to make a judgment and form a response (Wanke & Schwarz, 1997). Contrast effects may occur in a couple of ways at the judgment and response stages. When items on a survey are redundant in content, some respondents may wrongly assume that the content in each item is new and intentionally respond differently to the redundant item(s), resulting in contrast effects (Tourangeau & Rasinski, 1988; Wanke & Schwarz, 1997). Additionally, some respondents may want to provide "fair" responses, demonstrating that they have considered both sides of an issue, which can also cause contrast effects (Wanke & Schwarz, 1997).

Item order effects may be less problematic in a self-administered survey than in an orally administered survey. When respondents complete a self-administered survey, they have more time to review their responses to prior items. There is less pressure to respond quickly because there is no survey administrator waiting for a response (Schwarz, Strack, Hippler, & Bishop, 1991).

ii. Response order

The order of the response options may also affect how a respondent selects an option (Galesic, Tourangeau, Couper, & Conrad, 2007; Krosnick, 1991, 1999). Using eye-tracking data, Galesic and colleagues verified Krosnick's suspicion that respondents spend more time reading and interpreting the first response options that they see in survey items in comparison to the latter options; this effect occurred irrespective of the content of the response options. This phenomenon may occur because of *primacy effects* (i.e., a tendency to process

initial response options more thoroughly than latter options) (Galesic et al., 2007). Another explanation for respondents spending more time on the first response options may be that respondents do not bother to look at the latter options (Galesic et al., 2007; Krosnick, 1991). In his theory of satisficing, Krosnick (1991) proposed the selection of the first plausible response option as a shortcut to reduce the cognitive load on the respondent. This effect intensifies as the number of response options increases, requiring respondents to exert ever-increasing cognitive effort to read, interpret, and select the most appropriate responses (Galesic et al., 2007; Krosnick, 1991). These findings provide further support for the recommendation to use the smallest number of response options necessary in order to encompass fully all meaningful divisions of the concept of inquiry.

b. Visual design elements

Cognitive processing demands may be greater for self-administered surveys because they include elements of visual perception (Jenkins & Dillman, 1997). Even small changes to the visual design of a survey instrument may affect responses to items and encourage or discourage satisficing behavior (Galesic et al., 2007). Visual elements can reduce cognitive load by making a survey instrument easier to read and interpret. However, respondents have a tendency to consider every visual element of a survey instrument significant when interpreting and responding to items (Tourangeau et al., 2007). For this reason, it is important to consider the various visual design elements of a survey instrument.

When determining how to respond to a particular item, respondents must distinguish between different types of survey elements -- those that are *task elements*, or vital to interpreting and responding to the item, and those that are *stylistic elements*, or secondary features of the survey instrument that are not intended to convey meaning (i.e., the respondent does not need to

pay attention to these elements in order to respond to items) (Couper et al., 2004). Examples of stylistic elements may include font type and color, spacing, and background color (Couper et al., 2004).

Regardless of the intention of the survey designer, respondents may inappropriately attend to some stylistic elements when they are responding to the item(s). In these cases, respondents react differently when certain style elements are present than they do when they are not present.

Couper et al. (2004) warned that certain stylistic elements may also add to cognitive load. For instance, font that is difficult to read, or background and font color combinations that do not provide sufficient contrast for respondents to easily read items, may increase the cognitive effort required to respond to a survey instrument. For example, Tourangeau et al. (2007) found that shading had a small, but detectable effect on responses to scaled favor/oppose items, shifting responses slightly to the more favorable end of the scale. This shading effect was very small, and other cues, such as descriptive labels, seemed to eliminate the effect. Tourangeau et al. (2007) hypothesized that there may be a hierarchy of cues that respondents attend to when searching for information regarding how to respond to an item (i.e., descriptive labels take precedence over visual-only information, such as color).

In web-based surveys, how a survey designer chooses to present response options may also affect the responses (Galesic et al., 2007). Results from eye-tracking studies suggest that some respondents tend not to read and consider response options that are displayed in a drop-down box (Galesic et al., 2007). In this format, response options are not immediately visible. The respondent must click on the drop-down box in order to see the entire list of options. As an

alternative, Galesic and his colleagues recommended that survey designers use radio buttons, since in this format all the response options are immediately visible.

At times, survey designers may deem it necessary to include one or more important definitions that respondents should use when interpreting an item. If survey designers want to use definitions in web-based items, it may be best to present the definitions on-screen in a manner that does not require extra clicks or hovering the mouse pointer over the word(s) to access those definitions, as even this small amount of additional cognitive effort seems to reduce the likelihood that respondents will take the time to access the definitions (Galesic et al., 2007).

H. Comparison of Two Theoretical Models

The question-answering model that Cannell (1981) and his colleagues proposed and Tourangeau's (1984) model explaining the role of cognition in survey response have many similarities (see Figure 1). Both models include the same basic cognitive steps: (a) understanding/interpreting the item, (b) retrieving relevant information, (c) using that information to make a judgment or evaluation, and (d) responding to the item.

The biggest difference between the two models is that the model that Cannell and his colleagues created includes the potential for an additional step in Steps 1, 2, or 3 (interpretation, retrieval, and judgment) of Tourangeau's model. In that additional step, respondents may consider irrelevant cues in the survey or environment. If respondents consider such cues, then they may make an inadequate response to an item. Tourangeau's model does not explicitly take the role of irrelevant cues into account.

The two models also differ in terms of how they conceptualize the judgment (or evaluation) process that respondents use. Cannell and his colleagues view rendering a judgment or evaluation as a two-step process (Step 3: Evaluation of response accuracy, and Step 4:

Evaluation of response based on other goals (e.g., expressing socially desirable opinions)). By contrast, Tourangeau included the making of a judgment or evaluation as a single step in his model without separating details regarding how respondents might evaluate their responses into separate steps.

Finally, while both models include a response step, Cannell and his colleagues differentiate between respondents making an appropriate response to an item and respondents making an inadequate response. For these theorists, respondents may make inadequate responses in Steps 2, 3 or 4 if they modify their judgments or evaluations based on extraneous cues.

I used Tourangeau's (1984) model as the basis for the theoretical model that I created for reducing cognitive load in survey instruments. While there is no empirical evidence to suggest that Tourangeau's model is superior to Cannell's, there were two reasons for my choosing Tourangeau's model. First, one of the main goals of my studies was to design a simple and succinct theoretical model for practitioners to use to guide survey design. Tourangeau's model for the cognitive analysis of the task of the respondent is simpler and more succinct than the model that Cannell and his colleagues created to characterize a respondent's question-answering process. Second, in my review of the literature, it seems that researchers have more frequently used Tourangeau's model; thus, it may have more "currency" in the survey design community than the model that Cannell and his colleagues proposed.

I. Proposing a Theoretical Model for Reducing Cognitive Load

The theoretical model for reducing cognitive load that I am proposing reflects a series of best practices for survey design. It is based on key findings from the cognitive load research literature. I have organized the best practices into a series of guidelines for reducing the cognitive load of surveys. Additionally, some of the guidelines have specific recommendations that

accompany them. Research findings suggest that implementation of these best practices should reduce the amount of cognitive effort required to respond to survey items and instruments. When survey designers use these best practices, it should result in lower cognitive load for the respondents and better survey outcomes.

My theoretical model is aligned with the Tourangeau (1984) model (i.e., there are guidelines in my theoretical model that pertain to each of the four steps in Tourangeau's model for explaining the role of cognition in survey response). Figure 2 shows the organizational framework for my proposed theoretical model.

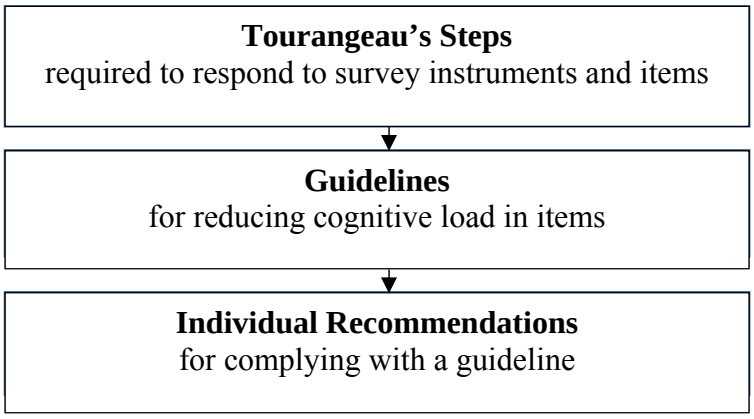

Figure 2. Organizational framework for the proposed theoretical model.

In the outline that follows, I present my proposed theoretical model for reducing cognitive load.

Tourangeau's Step 1: Interpretation

Item Stems:

1. **Guideline**: Use language that is clear and familiar to survey respondents.
    a. **Recommendation**: Avoid cognitively taxing wording.
    b. **Recommendation**: Avoid unfamiliar words and phrasing.
    c. **Recommendation**: Avoid *jargon* (i.e., words or phrases that may be familiar to certain groups or profession, but are unfamiliar to the general public) and *acronyms* (i.e., shortening a name or phrase by replacing words with single letters).
2. **Guideline**: Ensure that item stems are clear and explicit.
    a. **Recommendation**: Do not use concepts that are unclear or unfamiliar to respondents.
    b. **Recommendation**: Avoid complex sentence structures.
    c. **Recommendation**: Ask about only one concept in each stem; avoid *double-barreled* items (i.e., asking about more than one concept in a single item).
3. **Guideline**: Use item stems that do not make assumptions.
4. **Guideline**: Ask for information in a direct manner by avoiding *double negatives* (i.e., requiring a negative response to a negatively phrased item stem in order to provide a positive response to the item).
5. **Guideline**: Ensure item stems are succinct, including only as much information as is necessary for respondents to properly interpret what is being requested of them.

Survey Construction:

1. **Guideline**: Include clear instructions that clarify the purpose of the survey instrument, and provide respondents with expected procedures for responding to the survey instrument.
2. **Guideline**: Ensure that every portion of a survey instrument is visible without the need for additional action by the respondent.
    a. **Recommendation**: Use *radio buttons* (i.e., bubbles that appear on the screen next to options) instead of *drop-down boxes* (i.e., a list of options that a respondent observes after clicking on a down arrow in the survey instrument) to display response options.
    b. **Recommendation**: Do not "hide" definitions respondents may need to interpret and respond to survey items.
3. **Guideline**: Use easy-to-read font size and type.
4. **Guideline**: Use high-contrast font and background colors.

Tourangeau's Step 2: Retrieval

Item Stems:

1. **Guideline**: Use stems that request information with which respondents have primary experience and avoid asking for *second-hand information* (i.e., information that the respondent has heard about, but not experienced personally) or hypothetical information.

Response Options:

1. **Guideline**: Group conceptually similar items together.

Tourangeau's Step 3: Judgment

Response Options:

1. **Guideline**: Use the smallest number of response options necessary to encompass all meaningful divisions of the *concept of inquiry* (i.e., the topic for which a survey is requesting information).
2. **Guideline**: Generally, use four response options if not including a neutral option and five response options if a neutral option will be included.
3. **Guideline**: Neutral response options should only be included if a survey designer reasonably expects some respondents to have no opinion; otherwise, avoid including neutral options.

Tourangeau's Step 4: Response

Response Options:

1. **Guideline**: Only label the most *extreme response options* (i.e., the highest and lowest options) on a scale.
2. **Guideline**: When possible, use the same scaled response options for similar items in a survey instrument.
3. **Guideline**: If using numeric scale labels, use only positive numbers.
4. **Guideline**: Use the smallest number of response options necessary to encompass all meaningful divisions of the concept of inquiry

III. METHOD

To answer my research questions, I carried out two primary studies after I completed an initial pilot study. Analysis of the survey data I obtained from Primary Study 1 helped me answer my first research question, while the results from Primary Study 2 tested the efficacy of my theoretical model and helped me answer my final four research questions.

In order to test the efficacy of my theoretical model, I measured cognitive load using a set of survey items that met the guidelines specified in my theoretical model and a set of items that were very similar in content but did not meet those guidelines. While there are multiple ways to collect information about the cognitive load of survey items (described in the cognitive load theory and measurement section of this dissertation), I used two of the three methods that researchers (e.g., Paas, Tuovinen et al., 2003; Paas & Van Merrienboer, 1994b) have found effective. In Primary Study 1, I presented students with two versions of 22 items. The two versions of each item measured the same content but differed in terms of their cognitive load. One was a high cognitive load (HCL) version, and the other was a low cognitive load (LCL) version. I asked the student to read the two versions and then decide which one required more mental effort to respond. In Primary Study 2, I randomly assigned students to complete a survey containing all HCL items or a survey containing all LCL items and then recorded the amount of time that each student spent reading and answering each item on the survey.

Some researchers argue that intensity of effort while performing a task is the most reliable estimate of cognitive load (Paas, Tuovinen et al., 2003). In Primary Study 2, I used time on task (or the amount of time taken to read and respond to a survey item) as a proxy measure for intensity of effort. However, because the instruments were administered online, I could not determine if a high response time for an item meant that a college student expended much mental

effort thinking about that item, or if the student may have walked away from the computer, answered a phone call, decided to check email, etc. Alternatively, a low response time for an item might have indicated high cognitive load if a student chose to skip an item or use a response set when responding to it. Unfortunately, the interpretation of the meaning of low and high response times is not straightforward. Therefore, I would need to exercise caution when interpreting any differences I found in the times that students took to respond to parallel versions of an item. For this reason, I used a second different, but complementary, method for measuring cognitive load (i.e., asking students to decide which of the two versions of each item required more mental effort to process).

The third method that researchers employ when measuring cognitive load involves the use of physiological techniques. Examples of physiological measures are eye tracking, heart rate, and blood pressure. I chose not to use physiological measures because of their intrusive nature and potential lack of sensitivity to subtle differences in cognitive load (Paas & Van Merrienboer, 1994b).

In the next section of this chapter, I describe the eligibility guidelines I employed when selecting my student samples. I then describe the pilot study that I conducted. After I completed the pilot study, I carried out two primary studies using different student samples. I describe the demographic characteristics of each of those samples and the recruitment strategies I employed. At the end of this chapter, I discuss the methods I used to collect data for Primary Study 1 and Primary Study 2.

A. Eligibility Guidelines for Selecting Student Samples

Although survey design is important in many different fields, I chose college students as my respondent group because of my particular interest in surveys administered for the purposes of gathering information about the assessment of student learning in higher education. In these types of surveys, college students are frequently the target group for response. Because I needed to obtain sufficiently large samples to draw conclusions about the cognitive load of survey items and the effects of cognitive load on survey response, I did not put many limits on eligibility to participate in my research. The participants needed to be college students, though they could be undergraduate or graduate, and they could be enrolled in any institution of higher education. They also needed to be at least 18 years old to ensure that they were able to consent to participation in the research.

Although I did not conduct any subgroup analyses of my data, it was important to document the demographic characteristics of my samples because they were convenience samples. In Table I, I describe the demographic information that I solicited from the students who participated in my studies.

B. Pilot Study

I created two sets of items to measure students' satisfaction with various aspects of their college experience. The two versions of each item were similar in content, but they differed in their theoretical cognitive loads: one version in each set was an item with theoretically low cognitive load (LCL), and the other version in each set was an item with theoretically high cognitive load (HCL). Each item presented a statement, and students were to indicate the extent to which they agreed or disagreed with that statement. All items used the same 4-point agreement response scale: strongly agree, agree, disagree, or strongly disagree. To accommodate those

TABLE I
DEMOGRAPHIC INFORMATION SOLICITED FROM STUDENTS

Demographic Characteristic	**Question**	**Response Options**
Gender	Please indicate your gender.	Male, Female, Other
Age	Please type the year in which you were born* *note, you must be at least 18 years old to participate in this survey	Open response
Race/Ethnicity	Please select the option that best describes your race/ethnicity (select all that apply)	American Indian or Alaska Native; Asian; Black or African American; Hispanic or Latino; Native Hawaiian or Pacific Islander; White
Institution Type	Please select the option that best describes the type of higher education institution you attend.	2 year/community college; 4 year public college or university; 4 year private college or university; Other
Education Level	What is the highest level of education you have completed?	High school diploma or GED, Associate's degree or equivalent, Bachelor's degree or equivalent, Master's degree or equivalent, Doctoral or equivalent professional degree, Other (if selected, an open-ended item appeared asking to describe)
Education Student is Seeking	Please indicate the type of education you are currently seeking	Associate's degree; Bachelor's degree; Master's degree; Doctoral degree or equivalent professional degree; Other (if selected, an open-ended item appeared asking to describe)

students who may have lacked exposure to (or may have been unfamiliar with) certain aspects of the college experience described in one (or more) of the items, I included a fifth, unscaled, option of "not applicable."

I tested the efficacy of only a portion of my model (i.e., the interpretation portion), not the entire model, since to test the entire model would have required the design and administration of many more items than could be included in a single research study. Although there are nine guidelines in the interpretation portion of the model with 14 specific recommendations, I tested eight of these guidelines and 12 specific recommendations that I could manipulate (see Table II). To increase the reliability of the final survey instrument, I included two item pairs for each recommendation.

The purpose of the pilot study was to determine to what extent a small group of college students (n = 10) agreed that the HCL versions of items that I created required more cognitive effort to respond to than the LCL versions of those same items. The goals of the pilot study were to (a) identify versions of items that may not have functioned as I intended them to function (i.e., as HCL or LCL versions), (b) identify versions of items that the students did not understand, and (c) pinpoint the sources of their confusion and/or misunderstanding. Conducting this pilot study gave me the opportunity to revise items prior to assembling my survey instruments for my two primary studies.

For the pilot study, I recruited college students using local listservs. I sent a recruitment email to a local community of practice on April 8 to recruit participants. I forwarded the recruitment email to a few individuals who were members of the community of practice because they personally contacted me and asked me to send them the recruitment email. The recruitment period for the pilot study was from April 8–28, 2016. The research participants were either

TABLE II

MODEL GUIDELINES AND RECOMMENDATIONS TESTED
USING LCL AND HCL ITEMS

Model Guideline	**Specific Recommendations**	**LCL Items**	**HCL Items**
Use language that is clear and familiar to survey respondents.	Avoid cognitively taxing wording.	I am satisfied with my advisor's ability to address my needs	I am satisfied with my advisor's aptitude in addressing my needs.
		My institution communicates effectively with students about safety issues.	My institution effectively transmits to students information regarding safety issues.
	Avoid unfamiliar words and phrasing.	I am satisfied with the academic challenge of the courses at my institution.	I am satisfied with the academic rigor of the courses at my institution.
		My instructors provide me with feedback that helps me improve.	My instructors provide me with formative feedback.
Use language that is clear and familiar to survey respondents. *continued*	Avoid jargon and acronyms.	I can easily navigate my institution's learning management system (for example, Blackboard, Canvas, Desire2Learn, Sakai, Moodle).	I can easily navigate my institution's LMS.
		My institution has sufficient information available about applying for federal financial aid.	My institution has sufficient information available about completing the FAFSA.

TABLE II *continued*

Model Guideline	**Specific Recommendations**	**LCL Items**	**HCL Items**
Ensure item stems are clear and explicit.	Do not use concepts that are unfamiliar or unclear to respondents.	My institution provides students with sufficient information about what to do in case of an emergency.	My institution provides students with sufficient information about its emergency preparedness plan.
		I am satisfied with the assistance that my institution provides to support my computer usage.	I am satisfied with the infrastructure support that my institution provides to support my computer usage.
	Avoid complex sentence structure.	I am satisfied with the access I have to academic journals at my institution.	Considering the access to academic journals I have at my institution, I am satisfied.
		My instructors have good knowledge about the content they teach.	Based on my experiences with instructors with whom I have taken classes, instructors at my institution have good knowledge about the content they teach.

TABLE II *continued*

Model Guideline	**Specific Recommendations**	**LCL Items**	**HCL Items**
Ensure item stems are clear and explicit. *continued*	Ask about only one concept in each stem; avoid double-barreled items	I am satisfied with: the variety of elective (i.e., non-required) courses at my institution. the availability of elective (i.e., non-required) courses at my institution (Phrasing as a two-part item so that the student responds to only one concept in each part).	I am satisfied with the variety and availability of elective (i.e., non-required) courses at my institution.
		I am satisfied with: the types of activities available for me to participate in at my institution. the times that activities are scheduled at my institution. (Phrasing as a two-part item so that the student responds to only one concept in each part).	I am satisfied with the types of activities available for me to participate in at my institution and the times that those activities are scheduled.

TABLE II *continued*

Model Guideline	**Specific Recommendations**	**LCL Items**	**HCL Items**
Use item stems that do not make assumptions	Use item stems that do not make assumptions	I am satisfied with the availability of career planning services at my institution.	I am satisfied with the career planning services I have received at my institution.
		I am satisfied with the process to register for classes at my institution.	I am satisfied with the different options to register for classes at my institution.
Ask for information in a direct manner by avoiding double negatives.	Ask for information in a direct manner by avoiding double negatives.	I am satisfied with the types of assignments required in my classes.	I am not satisfied with the types of assignments required in my classes.
		I am satisfied with the diversity of perspectives that I have been exposed to at my institution.	I am not satisfied with the diversity of perspectives that I have been exposed to at my institution.
Ensure that item stems are succinct, including only as much information as is necessary for respondents to properly interpret what is being requested of them.	Ensure that item stems are succinct, including only as much information as is necessary for respondents to properly interpret what is being requested of them.	I am satisfied with the internet connectivity on campus.	Thinking about all the ways I use the internet, including web browsing, email, and social media, I am satisfied with the internet connectivity on campus.
		I am satisfied with the opportunities for community service at my institution.	Thinking about all the various opportunities that my institution makes available for me to participate in community service, I am satisfied with the opportunities for community service at my institution.

TABLE II *continued*

Model Guideline	**Specific Recommendations**	**LCL Items**	**HCL Items**
Use clear instructions.	Use clear instructions.	See Table VII	See Table VII
Ensure that every portion of a survey instrument is visible without the need for additional action by the respondent.	Use radio buttons instead of drop-down boxes to display response options.	My degree will prepare me well for the career path that I would like to pursue. (respond using radio buttons)	My degree will prepare me well for the career path that I would like to pursue. (respond using drop-down box)
		The other students in my courses help me to learn course material better. (respond using radio buttons)	The other students in my courses help me to learn course material better. (respond using drop-down box)
	Do not "hide" definitions respondents may need to interpret and respond to survey items.	There is no functionality to hide definitions in the version of Qualtrics© available to me.	
Use easy-to-read font size and type.	Use easy-to-read font size and type.	There is a strong sense of school spirit at my institution.	*There is a strong sense of school spirit at my institution.*
		My instructors show concern for how much I am learning in my classes.	*My instructors show concern for how much I am learning in my classes.*
Use high-contrast font and background colors.	Use high-contrast font and background colors.	There is no functionality to change background or font colors in the version of Qualtrics© available to me.	

students at member institutions of the community of practice or students who were affiliated in some way with faculty or staff at a member institution. All students were local. This was necessary because students needed to be available to meet in person during the study period.

I gave each student a $15 Amazon.com gift card as an incentive for participating. Ten students participated in my pilot study. I did not collect demographic information from these students. However, since I conducted the pilot study via in-person interviews, I do know that nine of the students were female, and one was male.

I met with each student individually. At the beginning of the session, I gave the student an informed consent document and asked him/her to read it and sign it. Once I obtained informed consent, I presented both the HCL and LCL versions of each item side-by-side on a computer screen using Qualtrics© software. (See Appendix A.) After the student read both versions, I asked which one seemed to require more mental effort, or 'thinking,' in order to provide a response. I defined *mental effort* as "how much you had to think about the statement to give a response. For example, if a statement requires more mental effort to understand, you might have to read it twice." Once the student selected a version, I asked him/her to tell me why that particular version seemed more cognitively demanding. I also asked him/her whether each version was clear, and if not, what part(s) of the statement(s) was/were confusing, ambiguous, and/or hard to understand. I audio recorded each student's responses and then listened to the audiotapes to take notes.

I conducted a content analysis of my notes to identify versions of items that did not perform as intended and document the reasons why (e.g., there was not a clear distinction between the HCL and LCL versions of an item in terms of the amount of mental effort required to respond, or there were LCL versions of items that some students believed required more

mental effort to respond to than the HCL versions). I also identified items that needed revision for clarity and used the results from this initial pilot study to revise problematic versions of items.

C. Demographic Characteristics of the Student Sample Recruited for Primary Study 1

To recruit students for Primary Study 1, I sent a recruitment email to a community of practice listserv for a large midwestern urban area. I also sent a recruitment email to a large midwestern university's graduate program's student listserv. Because I administered the survey for Primary Study 1 online, it was not necessary that students who wanted to participate in this study live in the area. This allowed me to use a national recruitment strategy: I sent an email to a national professional listerv to recruit additional students. The recruitment period for Primary Study 1 was May 31 – June 30, 2016. Sixty-four students participated in Primary Study 1. I collected demographic information from these students (see Table III).

D. Demographic Characteristics of the Student Sample Recruited for Primary Study 2

For Primary Study 2, I needed to recruit a study sample that did not include any students who were involved in my pilot study or in Primary Study 1. Therefore, I used different methods for recruiting students to participate in this study. While I cannot be certain that there were no students who participated in more than one of my studies, my having used a different recruitment strategy for each study should have at least mitigated this possibility. I used social media and a student listserv to recruit students for Primary Study 2. I posted my recruitment message on both Twitter and Facebook. In addition, I used a midwestern large urban university's mass email system to send a recruitment email to all university students who had active email addresses.

TABLE III

DEMOGRAPHIC INFORMATION REPORTED BY STUDENTS WHO PARTICIPATED IN PRIMARY STUDY 1

Demographic Characteristic	**Response Options**	**# of Students**	**% of Students**
Gender	Male	39	60.94%
	Female	22	34.38%
	Other	2	3.13%
	Missing Data	1	1.56%
Age	18-24	28	43.75%
	25-34	13	20.31%
	35 or older	11	17.19%
	Missing Data	12	18.75%
Race/Ethnicity[a]	American Indian or Alaska Native	2	3.13%
	Asian	3	4.69%
	Black or African American	1	1.56%
	Hispanic or Latino	3	4.69%
	Native Hawaiian or Pacific Islander	0	0.00%
	White	52	81.25%
	Missing Data	3	4.69%
Type of Institution Attending	Two Year/Community College	0	0.00%
	Four Year Public	37	57.81%
	Four Year Private	22	34.38%
	Other	0	0.00%
	Missing Data	5	7.81%
Education Level Completed	High school diploma or GED	14	21.88%
	Associate's degree or equivalent	4	6.25%
	Bachelor's degree or equivalent	25	39.06%
	Master's degree or equivalent	19	29.69%
	Doctoral degree or equivalent professional degree	1	1.56%
	Other	0	0.00%
	Missing Data	1	1.56%
Education Student is Seeking[a]	Associate's Degree	0	0.00%
	Bachelor's degree	18	28.13%
	Master's degree or equivalent	29	45.31%
	Doctoral degree or equivalent professional degree	16	25.00%
	Other	0	0.00%
	Missing Data	1	1.56%

[a] *Note*. Students were able to select multiple response options for race/ethnicity and for the educational degrees that the students were seeking.

The recruitment period for this phase of the research was June 27 – August 5, 2016.

I had initially planned to recruit at least 125 students for this study. However, the first 125 eligible students (i.e., those who did not ask to be excluded) who responded to the surveys did not use the response options evenly. Linacre (2002) recommended that for a polytomous response scale, the minimum sample size needed to provide stable item estimates is 25 * (the number of response options + one). Since my response scale had four options, I needed a minimum sample size of 125 students, according to this recommendation. Additionally, Linacre cautioned that if the students do not use the response options evenly, up to 100 * (the number of response options + one) may be necessary. Therefore, I needed to recruit additional students. While I did not impose an upper limit on the number of students who could participate in this study, I had hoped to recruit between 125 and 500 students.

Five hundred fifty-seven students participated in Primary Study 2; 277 students responded to the items on the Low Cognitive Load (LCL) instrument, and 280 students responded to the items on the High Cognitive Load (HCL) instrument. I collected the same demographic data for students who took part in Primary Study 2 as I did for students who took part in Primary Study 1 (see Tables IV and V).

E. Data Collection for Primary Study 1

After I revised the versions of the items based on the information I obtained from students who participated in the pilot study, I carried out Primary Study 1 using a larger sample of college students to determine whether, in their views, it required more mental effort to read and respond to the HCL versions of the items than to read and respond to the LCL versions.

TABLE IV

DEMOGRAPHIC INFORMATION REPORTED BY STUDENTS WHO PARTICIPATED IN PRIMARY STUDY 2—LOW COGNITIVE LOAD SURVEY INSTRUMENT

Demographic Characteristic	Response Options	# of Students	% of Students
Gender	Male	101	36.46%
	Female	175	63.18%
	Other	1	0.36%
	Missing Data	0	0.00%
Age	18-24	154	55.60%
	25-34	87	31.41%
	35 and up	30	10.83%
	Missing Data	6	2.17%
Race/Ethnicity[a]	American Indian or Alaska Native	3	1.08%
	Asian	70	25.27%
	Black or African American	22	7.94%
	Hispanic or Latino	48	17.33%
	Native Hawaiian or Pacific Islander	3	1.08%
	White	140	50.54%
	Missing Data	0	0.00%
Type of Institution Attending	Two Year/Community College	10	3.61%
	Four Year Public	223	80.51%
	Four Year Private	27	9.75%
	Other	16	5.78%
	Missing Data	1	0.36%
Education Level Completed	High school diploma or GED	86	31.05%
	Associate's degree or equivalent	39	14.08%
	Bachelor's degree or equivalent	97	35.02%
	Master's degree or equivalent	50	18.05%
	Doctoral degree or equivalent professional degree	3	1.08%
	Other	1	0.36%
	Missing Data	1	0.36%
Education Student is Seeking[a]	Associate's Degree	3	1.08%
	Bachelor's degree	121	43.68%
	Master's degree or equivalent	71	25.63%
	Doctoral degree or equivalent professional degree	82	29.60%
	Other	6	2.17%
	Missing Data	0	0.00%

[a] *Note*. Students were able to select multiple response options for race/ethnicity and for the educational degrees that the students were seeking.

TABLE V

DEMOGRAPHIC INFORMATION REPORTED BY STUDENTS WHO PARTICIPATED IN PRIMARY STUDY 2—HIGH COGNITIVE LOAD SURVEY INSTRUMENT

Demographic Characteristic	**Response Options**	**# of Students**	**% of Students**
Gender	Male	99	35.36%
	Female	172	61.43%
	Other	1	0.36%
	Missing Data	8	2.86%
Age	18-24	148	52.86%
	25-34	87	31.07%
	35 and up	31	11.07%
	Missing Data	14	5.00%
Race/Ethnicity*	American Indian or Alaska Native	5	1.79%
	Asian	85	30.36%
	Black or African American	18	6.43%
	Hispanic or Latino	53	18.93%
	Native Hawaiian or Pacific Islander	1	0.36%
	White	137	48.93%
	Missing Data	0	0.00%
Type of Institution Attending	Two Year/Community College	11	3.93%
	Four Year Public	224	80.00%
	Four Year Private	23	8.21%
	Other	14	5.00%
	Missing Data	8	2.86%
Education Level Completed	High School or GED	81	28.93%
	Associate's degree or equivalent	38	13.57%
	Bachelor's degree	97	34.64%
	Master's degree or equivalent	46	16.43%
	Doctoral degree or equivalent professional degree	9	3.21%
	Other	2	0.71%
	Missing Data	7	2.50%
Education Student is Seeking*	Associate's Degree	3	1.07%
	Bachelor's degree	117	41.79%
	Master's degree or equivalent	74	26.43%
	Doctoral degree or equivalent professional degree	85	30.36%
	Other	6	2.14%
	Missing Data	0	0.00%

Note. Students were able to select multiple response options for race/ethnicity and for the educational degrees that the students were seeking.

I had planned to recruit at least 100 college students to participate in Primary Study 1. However, I encountered a problem with the initial administration of the online survey instrument: It appeared that at least half of the students that I had recruited never saw any of the items on the survey instrument. Those students only saw and responded to the initial request that they supply their demographic information. Unfortunately, neither Qualtrics© personnel nor I were able to determine that a problem had occurred. This unanticipated survey administration issue resulted in little useable data and meant that I needed to recruit additional students.

Ultimately, I collected data from 143 students for this study, but only 64 of those students provided responses to items on the survey instrument. However, this sample size provided statistical sensitivity to detect a 20% effect size. (I used G*Power software to calculate the sample size required to achieve 80% power on a comparison of proportions when one is fixed or known, resulting in a minimum sample size of 37.)

I chose to use Qualtrics©, an online survey system available to students, faculty, and staff at the university, to deliver the surveys in both primary studies because it offers more flexibility than the more commonly used survey systems available to me (such as Survey Monkey or Google Forms). Qualtrics© is also free for students, faculty, and staff and supported by the university's computing center. Students accessed the instrument using a link that I provided to them. Only students who had the survey's link had access to the instrument.

The first page of the survey explained how to access the informed consent document. Students saw the following text:

> Below, you will see a link to a document called an 'Informed Consent' document. This document provides valuable information about this research, including why you are being asked to participate, the purpose and procedures, potential risks and benefits, other

options to participation, whom to contact if you have questions or concerns, and your ability to withdraw from the research. Please read this document carefully and print a copy for yourself.

Qualtrics© provided a link to the embedded informed consent document. Because the text of my informed consent document was long, this was the only way that I could include my informed consent document in a Qualtrics© survey.

The next request that students saw was "Please make sure you have carefully read the document linked above and indicate whether or not you would like to provide your consent to participate in this research." Students had to click a bubble to select from two response options: (a) "I consent to participating in this study," or (b) "I do NOT consent to participating in this study." If students indicated they did not want to provide consent, they received a message thanking them for their time and providing contact information for the researchers and the university's office for the protection of research subjects. If students consented to participating in the research, they were directed to a request for demographic information.

After providing their demographic information, students were shown instructions that explained how to respond to the items on the survey instrument. The instructions included the same definition of mental effort as the one that I used in the pilot study. Although it is standard practice to collect demographic information at the end of a survey, I requested demographic information prior to survey administration. Students who responded to one or more of the initial items on the instrument but did not finish were an important group in this study, and I wanted to ensure that I had demographic information about this sample of students.

The instrument presented on the same screen both the HCL and the LCL versions of each item. (The instrument appears in Appendix B.) The item pairs were presented in a random order

to reduce the potential for order effects. After students read the two versions of an item, the survey prompted them to identify the version that, in their views, required more mental effort in order to provide a response. They made their selections by clicking on the version of the item that they felt was more cognitively demanding. I left the survey open for one month.

All students who participated in Primary Study 1 were eligible to enter a drawing for three $50 Amazon© gift cards. They had the opportunity to provide their contact information after they had completed the survey. If students wished to provide their contact information for the drawing, they received a link to a different Qualtrics© survey to collect that information. Using this approach for gathering that contact information separated their responses to the survey items from their contact information.

F. Data Collection for Primary Study 2

For my second primary study, I constructed two instruments. One instrument contained all the LCL items that I administered in Primary Study 1, while the second instrument contained all the HCL items from that study. (The instruments appear in Appendix C.) Each student in this study responded to items in just one of the instruments. The purposes of this study were to compare (a) the amounts of time that students spent responding to the items on the two different instruments (i.e., one that contained HCL items, and one that contained LCL items), (b) the response rates for the two versions of each item (i.e., the HCL version and the LCL version), (c) the survey completion rates for the two instruments, and (d) the students' use of response sets when they provided their responses to the items.

I used the survey instructions recommendation (see Table VI) in my theoretical model to create a HCL version of the instructions for the HCL instrument and a LCL version of the instructions for the LCL instrument. I used the Lexile© software to evaluate each version. The

HCL instructions have a Lexile score of 660L (i.e., the equivalent of a 5th grade reading level), and the LCL instructions have a Lexile score of 1070L (i.e., the equivalent of a 3rd grade reading level) (MetaMetrics, 2015).

TABLE VI

MODEL GUIDELINES AND RECOMMENDATIONS TESTED USING LCL AND HCL SURVEY INSTRUMENTS—INSTRUCTIONS

Model Guideline	Specific Recommendations	LCL Version	HCL Version
Include clear instructions that clarify the purpose of the survey instrument and provide respondents with expected procedures for responding to the survey instrument.	Include clear instructions that clarify the purpose of the survey instrument and provide respondents with expected procedures for responding to the survey instrument.	Please read each statement and indicate your level of agreement with that statement. This is a brief survey; please complete it in one sitting if possible.	After reading each statement, register your level of agreement or disagreement with that statement. This is a brief survey; please complete it in one sitting if possible.

Because there is no way in Qualtrics© to display response options when the two versions of an item are shown side-by-side, I was not able to test in either the pilot study or Primary Study 1 this particular recommendation: Use radio buttons instead of drop-down boxes to display response options. The only difference between the two versions in each pair of items that tested this recommendation was in how the response options were presented (i.e., the wording of the statement and the response options for both versions of the item in each item pair were identical). Consequently, if I had included these item pairs in the pilot study or in Primary Study 1, it was highly unlikely that I would have gained useful information that could have informed revision of these particular pairs of items. However, I did include these item pairs in the survey instruments

that I employed in Primary Study 2. That is, the survey instruments employed in the pilot study and in Primary Study 1 included 22 items (i.e., 11 item pairs), while each survey instrument employed in Primary Study 2 included 24 items.

The first page of each survey employed in Primary Study 2 explained how to access the informed consent document. Students saw the following text:

> Below, you will see a link to a document called an 'Informed Consent' document. This document provides valuable information about this research, including why you are being asked to participate, the purpose and procedures, potential risks and benefits, other options to participation, whom to contact if you have questions or concerns, and your ability to withdraw from the research. Please read this document carefully and print a copy for yourself.

Qualtrics© then displayed the link to the embedded informed consent document. The next request that students saw was "Please make sure you have carefully read the document linked above and indicate whether or not you would like to provide your consent to participate in this research." Students selected one of two response options: (a) "I consent to participating in this study," or (b) "I do NOT consent to participating in this study." If students indicated they did not want to provide consent, they received a message thanking them for their time and providing contact information for the researchers and the university's office for the protection of research subjects. If students consented to participating in the research, they were directed to a request to supply demographic information.

After supplying their demographic information, students were shown instructions that explained how to respond to the items on the survey instrument. Again, although it is standard practice to collect demographic information at the end of a survey, I requested demographic

information prior to survey administration. Students who responded to one or more of the initial items on the instrument but did not finish were an important group in this study, and I wanted to ensure that I had demographic information about this sample of students.

I told students in the instructions that the purpose of the study was to determine the experiences that are most important to students in terms of their overall satisfaction with institutions of higher education. If I had let students know that I was planning to look at how they responded to items, then that could have affected their responses in ways unrelated to the measurement of cognitive load. Therefore, it was necessary to not be entirely forthright when I described the true purpose of the study.

Qualtrics© randomly assigned students to respond to the items appearing on either the HCL instrument or the LCL instrument. The instrument presented the items one at a time in a random order. After students read an item, they provided their responses by clicking on the bubble that was next to the response option they chose. Qualtrics© recorded the amount of time that the student spent reading and responding to each item. I left the survey open for one month.

After a student responded to the final item on the survey instrument, I notified him/her of the true purpose of the research and revealed the version (HCL or LCL) of the survey that the student had completed. I then explained why I did not disclose the true purpose of the study at the beginning of the survey and gave the student an opportunity to remove his/her data from the research. I made it clear that if the student withdrew consent, the student was still eligible to participate in the raffle for the IPad Mini and that withdrawing consent did not affect the student's odds of winning the raffle. Four students withdrew their consent based on this information; consequently, I did not include in my data set their demographic information or any of their responses to items.

All students who participated in Primary Study 2 were eligible to enter a drawing for an IPad Mini. They had the opportunity to provide their contact information after they had completed the survey. If students wished to provide their contact information for the drawing, they received a link to a different Qualtrics© survey to collect that information. Using this approach for gathering that contact information separated their responses to the survey items from their contact information.

IV. RESULTS

In this chapter, I will present the results of my research. For each of the three studies that I carried out (pilot study, Primary Study 1, and Primary Study 2), I will begin by discussing the statistical methods I used to analyze the data I received. I will then present the results for each of my studies, including effect sizes.

A. Pilot Study

In the pilot study, I collected data in person so that I could ask students to explain *why* they believed that the version of the item they were selecting in each item pair required more cognitive effort (or thinking) than the alternate version of the item. I tape recorded each of the individual student sessions and took notes based on these recordings. The purpose of the pilot study was to determine if the item pairs were functioning as I had intended (i.e., to determine if students would report that the High Cognitive Load (HCL) version of the item in each item pair required more thinking to provide a response than the Low Cognitive Load (LCL) version of that same item). Additionally, if any of the item pairs were not functioning as I had intended, I wanted to gather information that would help me revise those item pairs prior to conducting my primary studies.

For the most part, the item pairs functioned as I had anticipated; overall, about 83% of the time the students selected the HCL versions of the items that I had created, believing that they required more cognitive effort to provide responses than the LCL versions. However, there were a few item pairs that did not function as I had intended. I used the study's results and the students' rationales for their choices to determine which item pairs to keep and to identify those item pairs that needed to be revised (or dropped) (see Table VII).

TABLE VII

ITEM-BY-ITEM RESULTS FROM THE PILOT STUDY

LCL Item	HCL Item	# Selected HCL	% Selected HCL	Decision	Rationale
I am satisfied with my advisor's ability to address my needs	I am satisfied with my advisor's aptitude in addressing my needs.	10	100%	Keep Items	Item pair functioned as intended.
My institution communicates effectively with students about safety issues.	My institution effectively transmits to students information regarding safety issues.	8[a]	89%[a]	Keep Items	One student believed that the LCL version of the item was more cognitively demanding, reasoning that the HCL version was "shorter," but, in actuality, the HCL version was longer.
I am satisfied with the academic challenge of the courses at my institution.	I am satisfied with the academic rigor of the courses at my institution.	6	60%	Rewrite	Many students identified "rigor" and "challenge" as two distinct concepts. Several did not feel that one was necessarily more cognitively demanding than the other.
My instructors provide me with feedback that helps me improve.	My instructors provide me with formative feedback.	8	80%	Keep Items	"Formative feedback" seems to be an unfamiliar term for the majority of these students.

[a] One student did not provide a response for this item pair; however, the student did not request to skip it. I believe the student simply overlooked the item.

TABLE VII *continued*

LCL Item	HCL Item	# Selected HCL	% Selected HCL	Decision	Rationale
I can easily navigate my institution's learning management system (for example, Blackboard, Canvas, Desire2Learn, Sakai, Moodle).	I can easily navigate my institution's LMS.	8	80%	Keep Items	"LMS" seems to be an unfamiliar acronym for the majority of these students.
My institution has sufficient information available about applying for federal financial aid.	My institution has sufficient information available about completing the FAFSA.	2	20%	Rewrite Items	Since these students were very familiar with the FAFSA (and called it by its acronym), they noted that the LCL version was wordier and therefore, they believed that it was more cognitively demanding.
I am satisfied with the assistance that my institution provides to support my computer usage.	I am satisfied with the infrastructure support that my institution provides to support my computer usage.	9[a]	100%[a]	Keep Items	Item pair functioned as intended.
I am satisfied with the access I have to academic journals at my institution.	Considering the access to academic journals I have at my institution, I am satisfied.	10	100%	Keep Items	Item pair functioned as intended.

[a] One student did not provide a response for this item pair; however, the student did not request to skip it. I believe the student simply overlooked the item.

TABLE VII *continued*

LCL Item	HCL Item	# Selected HCL	% Selected HCL	Decision	Rationale
My instructors have good knowledge about the content they teach.	Based on my experiences with instructors with whom I have taken classes, instructors at my institution have good knowledge about the content they teach.	10	100%	Keep Items	Item pair functioned as intended.
I am satisfied with the availability of career planning services at my institution.	I am satisfied with the career planning services I have received at my institution.	7	70%	Rewrite Items	Only a few students noticed that an assumption was being made.
I am satisfied with the process to register for classes at my institution.	I am satisfied with the different options to register for classes at my institution.	7	70%	Rewrite Items	Only a few students noticed that an assumption was being made.
I am satisfied with the types of assignments required in my classes.	I am not satisfied with the types of assignments required in my classes.	9	90%	Keep Items	Item pair functioned as intended.
I am satisfied with the diversity of perspectives that I have been exposed to at my institution.	I am not satisfied with the diversity of perspectives that I have been exposed to at my institution.	9	90%	Keep Items	Item pair functioned as intended.

TABLE VII *continued*

LCL Item	HCL Item	# Selected HCL	% Selected HCL	Decision	Rationale
There is a strong sense of school spirit at my institution.	*There is a strong sense of school spirit at my institution.*	9	90%	Keep Items	Item pair functioned as intended.
My instructors show concern for how much I am learning in my classes.	*My instructors show concern for how much I am learning in my classes.*	9	90%	Keep Items	Item pair functioned as intended.
I am satisfied with the opportunities for community service at my institution.	Thinking about all the various opportunities that my institution makes available for me to participate in community service, I am satisfied with the opportunities for community service at my institution.	10	100%	Keep Items	Item pair functioned as intended.

Based on the information I gained from conducting the pilot study, I made the decision to rewrite five of the item pairs prior to conducting my primary studies. The original item pairs and the rewritten versions of those pairs are shown in Table VIII.

TABLE VIII

ITEM REVISIONS BASED ON THE RESULTS FROM THE PILOT STUDY

Original LCL	Original HCL	Rewritten LCL	Rewritten HCL
I am satisfied with the academic challenge of the courses at my institution.	I am satisfied with the academic rigor of the courses at my institution.	My institution's campus bookstore stocks an adequate assortment of textbooks.	My institution's campus bookstore stocks an adequate mélange of textbooks.
My institution has sufficient information available about applying for federal financial aid.	My institution has sufficient information available about completing the FAFSA.	I am satisfied with the cost of attendance at my institution.	I am satisfied with the COA at my institution.
I am satisfied with the availability of career planning services at my institution.	I am satisfied with the career planning services I have received at my institution.	I am satisfied with the availability of tutoring services at my institution.	I am satisfied with the tutoring services I have received at my institution.
I am satisfied with the process to register for classes at my institution.	I am satisfied with the different options to register for classes at my institution.	I am satisfied with the availability of resources to apply for on-campus jobs.	I am satisfied with the resources I have used to apply for on-campus jobs.
I am satisfied with the internet connectivity on campus.	Thinking about all the ways I use the internet, including web browsing, email, and social media, I am satisfied with the internet connectivity on campus.	I am satisfied with the internet connectivity on campus.	Thinking about all the ways I use the internet, I am satisfied with the internet connectivity on campus.

While it would be premature to make any strong claims based on the rationales that this small sample of students gave to support their choices (i.e., only ten students and predominantly female), I gained some valuable insights into their thinking about cognitive load in survey design that may warrant further study.

First, when reading the items, students frequently mentioned how common certain words or phrases were. They would notice if an item was stated in language that they used in their day-

to-day conversations. They would also notice if an item contained words or phrases with which they were not familiar. The students' comments seem to suggest that perhaps cognitive load needs to be considered differentially for different populations, taking into consideration how familiar the target population is with the concepts expressed in items and the language used to convey those concepts. That is, an item may represent high cognitive load for one population if they were not familiar with certain words or phrases that it contained, while that same item might represent lower cognitive load for another population who readily understood the item. In their efforts to reduce the cognitive load of items and instruments, survey designers may want to consider how familiar their target population will likely be with the words and phrases they are planning to use in their items.

I also learned in the pilot study that the recommendation about acronyms in my theoretical model may need to be modified. Because the students who participated in my study were very familiar with the acronym FAFSA, they felt that the version of the item that contained the acronym required less mental effort to respond to than the item that used the full name of the form. (A number of the students were not aware that FAFSA is the acronym for Free Application for Federal Student Aid.) This finding highlights the importance of taking into consideration the target population's familiarity with various acronyms when determining cognitive load. In this case, these college students were very familiar with the acronym FAFSA, but many did not know what the acronym stood for. For this target population, the version of the item containing the acronym may actually be less cognitively demanding than the version of the item that uses the full name of the form. However, other populations that are not familiar with FAFSA would likely find that the version of the item containing the acronym would be more cognitively

demanding than the version of the item containing the full name of the form, since they would not have previously encountered that acronym and thus would not know what it stood for.

All students participating in the pilot study indicated that they felt that items required less mental effort when they were short and succinct. Students frequently mentioned that items seemed "clearer" when they were succinct. Some students tended to choose the shorter versions of items as requiring less mental effort, even when those versions were designed to be the higher cognitive load versions.

Finally, I learned from the pilot study that using italics can change the meaning of items, but not in a consistent manner. Most students indicated that when they saw an italicized version of an item, they wondered why it was in italics and whether that item was somehow "more important" than the other items. Students tended to believe that every change to an element in a survey instrument was meaningful, and they searched for reasons to explain why some items were in italics, and some were not. One student indicated that when she saw an item in italics, it drew her attention to the item and signaled to her that she should read the item carefully.

B. Primary Study 1

I used the results from Primary Study 1 to answer **Research Question 1: Will students report that it takes more mental effort to respond to HCL items than to LCL items?**

The outcome measures needed to answer this question were, for each item, the proportion of students who selected the HCL version in each item pair, indicating that it required more mental effort to provide a response than the LCL version. For each item, I used a *z*-test to compare the proportion of students who selected the HCL version to the proportion expected if students were to respond randomly when making their selections (i.e., 0.50).

The formula for a *z*-test to compare proportions is given by:

$$z = (p_1 - p_2) / SE \quad (1)$$

where p_1 is the proportion of students who selected the HCL version of an item as requiring more mental effort to provide a response than the LCL version, p_2 is the test proportion (i.e., the proportion of students who would select the HCL version of an item if the students were to respond in a random manner), and SE is the standard error of the sample.

Overall, students selected most of the HCL versions of the items as requiring more mental effort to provide a response at a statistically significantly higher rate than would be expected if students were responding randomly ($p = .00$). Students selected 16 of the 18 HCL versions of the items at rates that were statistically significantly higher than if they had been responding in a random manner (see Table IX). For Item 2 and Item 12, students tended to select the HCL version more frequently than the LCL version, but not at rates that were statistically significantly higher than if they had been responding randomly. I then used the Holm-Bonferroni procedure to adjust the minimum alpha values to create stricter decision criteria for determining significance (Holm, 1979). Use of this statistical method corrects for the potential increase in Type I error due to running multiple *z*-tests on the same dataset. After applying this adjustment, I identified an additional three HCL versions of items that students tended to select more frequently than the LCL versions, but not at rates that were statistically significantly higher than if students had been responding randomly ($p = .00$) (i.e., Item 1, Item 7, Item 15).

I created two pairs of HCL and LCL items to test each of the recommendations in my proposed model. Table X presents the results from *z*-tests to compare the proportion of students who selected the HCL versions in the two pairs of items to the proportion expected if students were to respond randomly when making their selections (i.e., 0.50). For each item pair, the

TABLE IX

RESULTS FROM *Z*-TESTS TO COMPARE THE TEST PROPORTION TO THE PROPORTION OF STUDENTS WHO SELECTED THE HCL VERSION OF EACH ITEM

Item Text	**Item Selection**	***N***	**Observed Proportion**	**Test Proportion**	***z***	**Exact Sig. (2-tailed)[a]**
LCL: I am satisfied with the availability of tutoring services at my institution. HCL: I am satisfied with the tutoring services I have received at my institution.	I1 Selected HCL Selected LCL Total	 40 22 62	 .65 .35 1.00	.50	6.61	.03
LCL: I am satisfied with the availability of resources to apply for on-campus jobs. HCL: I am satisfied with the resources I have used to apply for on-campus jobs.	I2 Selected HCL Selected LCL Total	 36 27 63	 .57 .43 1.00	.50	3.25	.31
LCL: I am satisfied with the types of assignments required in my classes. HCL: I am not satisfied with the types of assignments required in my classes.	I3 Selected HCL Selected LCL Total	 47 17 64	 .73 .27 1.00	.50	10.67	.00
LCL: I am satisfied with the diversity of perspectives I've been exposed to at my institution. HCL: I am not satisfied with the diversity of perspectives I've been exposed to at my institution.	I4 Selected HCL Selected LCL Total	 50 14 64	 .78 .22 1.00	.50	10.67	.00
LCL: I am satisfied with the internet connectivity on my campus. HCL: Thinking about all the ways I use the internet, I am satisfied with the internet connectivity on my campus.	I5 Selected HCL Selected LCL Total	 53 11 64	 .83 .17 1.00	.50	12.81	.00

[a] These are adjusted alpha values obtained after implementing the Holm-Bonferroni (1979) procedure.

TABLE IX *continued*

Item Text	Item Selection	*N*	Observed Proportion	Test Proportion	*z*	Exact Sig. (2-tailed)[a]
LCL: I am satisfied with the opportunities for community service at my institution. HCL: Thinking about all the various opportunities that my institution makes available for me to participate in community service, I am satisfied with the opportunities for community service at my institution.	I6 Selected HCL Selected LCL Total	 54 10 64	 .84 .16 1.00	.50	15.65	.00
LCL: There is a strong sense of school spirit at my institution. HCL: *There is a strong sense of school spirit at my institution*.	I7 Selected HCL Selected LCL Total	 40 20 60	 .65 .35 1.00	.50	7.59	.03
LCL: My instructors show concern for how much I am learning in my classes. HCL: *My instructors show concern for how much I am learning in my classes.*	I8 Selected HCL Selected LCL Total	 43 18 61	 .70 .30 1.00	.50	9.33	.00
LCL: I am satisfied with my advisor's ability to address my needs. HCL: I am satisfied with aptitude in addressing my needs.	I9 Selected HCL Selected LCL Total	 48 16 64	 .75 .25 1.00	.50	11.39	.00
LCL: My institution communicates effectively with students about safety issues HCL: My institution effectively transmits to students information regarding safety issues.	I10 Selected HCL Selected LCL Total	 54 10 64	 .84 .16 1.00	.50	15.65	.00
LCL: My institution's campus bookstore stocks an adequate assortment of textbooks. HCL: My institution's campus bookstore stocks an adequate mélange of textbooks.	I11 Selected HCL Selected LCL Total	 55 9 64	 .86 .14 1.00	.50	16.37	.00

[a] These are adjusted alpha values obtained after implementing the Holm-Bonferroni (1979) procedure.

TABLE IX *continued*

Item Text	Item Selection	*N*	Observed Proportion	Test Proportion	*z*	Exact Sig. (2-tailed)[a]
LCL: My instructors provide me with feedback that helps me improve. HCL: My instructors provide me with formative feedback.	I12 Selected HCL Selected LCL Total	 39 25 64	 .61 .39 1.00	.50	4.98	.10
LCL: I can easily navigate my institution's learning management system (for example, Blackboard, Canvas, Sakai). HCL: I can easily navigate my institution's LMS	I13 Selected HCL Selected LCL Total	 46 17 63	 .73 .27 1.00	.50	10.48	.00
LCL: I am satisfied with the cost of attendance at my institution. HCL: I am satisfied with the COA at my institution.	I14 Selected HCL Selected LCL Total	 53 11 64	 .83 .17 1.00	.50	14.94	.00
LCL: My institution provides students with sufficient information about what to do in case of an emergency. HCL: My institution provides students with sufficient information about its emergency preparedness plan.	I15 Selected HCL Selected LCL Total	 41 24 65	 .63 .37 1.00	.50	5.96	.05
LCL: I am satisfied with the assistance my institution provides to support my computer usage. HCL: I am satisfied with the infrastructure support my institution provides to support my computer usage.	I16 Selected HCL Selected LCL Total	 53 11 64	 .83 .17 1.00	.50	14.94	.00
LCL: I am satisfied with the access I have to academic journals. HCL: Considering the access to academic journals I have at my institution, I am satisfied.	I17 Selected HCL Selected LCL Total	 53 10 63	 .84 .16 1.00	.50	15.54	.00

[a] These are adjusted alpha values obtained after implementing the Holm-Bonferroni (1979) procedure.

TABLE IX *continued*

Item Text	Item Selection	*N*	Observed Proportion	Test Proportion	*z*	Exact Sig. (2-tailed)[a]
LCL: My instructors have good knowledge about the content they teach. HCL: Based on my experiences within instructors with whom I have taken classes, instructors at my institution have good knowledge about the content they teach.	I18 Selected HCL Selected LCL Total	 52 12 64	 .81 .19 1.00	 .50	 14.23	 .00

[a] These is an adjusted alpha value obtained after implementing the Holm-Bonferroni (1979) procedure.

TABLE X

RESULTS FROM Z-TESTS TO COMPARE THE TEST PROPORTION TO THE PROPORTION OF STUDENTS WHO SELECTED THE HCL VERSIONS IN EACH ITEM PAIR

Recommendation	**Item Pair**	***N***	**Observed Proportion**	**Test Proportion**	***z***	**Exact Sig. (2-tailed)[a]**
Use item stems that do not make assumptions.	I1&I2 Selected HCL Selected LCL Total	 76 49 125	 .61 .39 1.00	.50	4.66	.02
Ask for information in a direct manner by avoiding double negatives.	I3&I4 Selected HCL Selected LCL Total	 97 31 128	 .76 .24 1.00	.50	11.11	.00
Ensure that item stems are succinct, including only as much information as is necessary for respondents to properly interpret what is being requested of them.	I5&I6 Selected HCL Selected LCL Total	 107 21 128	 .84 .16 1.00	.50	14.48	.00
Use easy-to-read font size and type.	I7&I8 Selected HCL Selected LCL Total	 94 30 124	 .76 .24 1.00	.50	11.18	.00
Avoid cognitively taxing wording.	I9 &I10 Selected HCL Selected LCL Total	 94 34 128	 .73 .27 1.00	.50	10.10	.00

[a] These are adjusted alpha values obtained after implementing the Holm-Bonferroni (1979) procedure.

TABLE X *continued*

Recommendation	Item Pair	*N*	Observed Proportion	Test Proportion	*z*	Exact Sig. (2-tailed)[a]
Avoid unfamiliar words and phrasing.	I11&I12					
	Selected HCL	119	.81	.50	13.34	.00
	Selected LCL	28	.19			
	Total	147	1.00			
Avoid jargon and acronyms.	I13&I14					
	Selected HCL	99	.78	.50	12.60	.00
	Selected LCL	28	.22			
	Total	127	1.00			
Do not use concepts that are unfamiliar or unclear to respondents.	I15&I16					
	Selected HCL	94	.73	.50	9.86	.00
	Selected LCL	35	.27			
	Total	129	1.00			
Avoid complex sentence structure.	I17&I18					
	Selected HCL	105	.83	.50	14.09	.00
	Selected LCL	22	.17			
	Total	127	1.00			

[a] These are adjusted alpha values obtained after implementing the Holm-Bonferroni (1979) procedure.

resulting *z* statistic was statistically significant. Applying the Holm-Bonferroni (1979) adjustment to alpha values for these item pairs did not alter these results.

I then calculated effect sizes for the *z*-test statistics I obtained to determine the magnitude of each effect. I used Psychometrica's effect size calculator (Lenhard & Lenhard, 2016) to estimate the effect size for the *z* statistic for each item and the *z* statistic for each item pair, and I interpreted the effect sizes using Cohen's (1988) recommendations.

As shown in Table XI, for 17 of the 18 items, the effect sizes indicated that the magnitude of each of those effects was strong, while for the remaining item (i.e., Item 2), the effect size indicated that the magnitude of that effect was moderate/intermediate. For all nine pairs of items, the effect sizes indicated that the magnitude of each of those effects was also strong (see Table XII).

C. Primary Study 2

Primary Study 2 involved the administration of an online survey. Students who participated in this study were randomly assigned to complete either a Low Cognitive Load (LCL) survey containing all the LCL versions of the items, or a High Cognitive Load (HCL) survey containing all the HCL versions of the items. Regardless of the version of the survey to which students were randomly assigned, they received items in a random order to reduce the potential for order effects. Based on their responses to the two surveys, I calculated the response rate for each survey and for each item on a survey, the time students took to respond to each of the surveys, and Rasch student fit statistics and point-measure correlations to detect response sets in the students' ratings.

TABLE XI

EFFECT SIZES FOR THE ITEMS

Item	***r* value**	**95% Confidence Interval**		**Interpretation**
		Lower	Upper	
I1	0.84	0.56	1.10	Strong Effect
I2	0.41	0.15	0.67	Intermediate Effect
I3	1.33	1.02	1.64	Strong Effect
I4	1.60	1.27	1.93	Strong Effect
I5	1.87	1.50	2.24	Strong Effect
I6	1.96	1.57	2.35	Strong Effect
I7	0.96	0.67	1.25	Strong Effect
I8	1.95	1.62	2.28	Strong Effect
I9	1.42	1.11	1.73	Strong Effect
I10	1.96	1.57	2.35	Strong Effect
I11	2.05	1.65	2.45	Strong Effect
I12	0.62	0.36	0.88	Strong Effect
I13	1.32	1.01	1.62	Strong Effect
I14	1.87	1.50	2.24	Strong Effect
I15	0.74	0.47	1.01	Strong Effect
I16	1.87	1.50	2.24	Strong Effect
I17	1.83	1.45	2.21	Strong Effect
I18	1.78	1.42	2.14	Strong Effect

TABLE XII

EFFECT SIZES FOR THE ITEM PAIRS

Item Pair	***r* value**	**95% Confidence Interval**		**Interpretation**
		Lower	Upper	
I1 & I2	0.92	0.73	1.10	Strong Effect
I3 & I4	0.98	0.76	1.20	Strong Effect
I5 & I6	1.28	1.03	1.53	Strong Effect
I7 & I8	0.73	0.52	0.94	Strong Effect
I9 & I10	1.06	0.85	1.27	Strong Effect
I11 & I12	0.89	0.67	1.11	Strong Effect
I13 & I14	1.10	0.88	1.32	Strong Effect
I15 & I16	0.87	0.66	1.08	Strong Effect
I17 & I18	1.25	1.00	1.50	Strong Effect

Research Question 2: Will more students complete the LCL instrument than the HCL instrument?

I used a *z*-test to compare the proportion of students who did not respond to any items on the LCL survey to the proportion of students who did not respond to any items on the HCL survey to determine if those proportions were statistically significantly different (i.e., were the response rates for the two surveys statistically significantly different). The equation for calculating this statistic is:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_0} \quad (2)$$

where $\hat{p}_1$ is the pooled (or weighted average) proportion of items for which students provided a response on the LCL survey and $\hat{p}_2$ is the pooled proportion of items for which students provided a response on the HCL survey, and SE_0 is the standard error of the difference between the two proportions if the null hypothesis were true (i.e., there is no difference between the response rates for the LCL and HCL surveys).

The results from this analysis are reported in Table XIII. Although there were relatively few students who did not respond to any of the items on the surveys, the proportion of students who did not respond to any of the items on the HCL survey was statistically significantly higher (p = .00) than the proportion of students who did not respond to any of the items on the LCL survey, though the effect size associated with this *z* statistic was small (r = .18), 95% CI [0.10, 0.27], indicating that the magnitude of the effect was weak.

Research Question 3: Will students respond to more items on the LCL instrument than to items on the HCL instrument?

I also used a *z*-test to determine whether the proportions of items that students skipped on the LCL and HCL surveys were statistically significantly different. As shown in Table XIX,

students responded to a significantly higher proportion of items on the LCL survey than on the HCL instrument (p =.00). However, the effect size associated with this z statistic was very small (r = .08), 95% CI [0.06, 0.10], indicating that the magnitude of the effect was weak.

TABLE XIII

RESULTS FROM A *Z*-TEST TO COMPARE THE PROPORTION OF STUDENTS WHO DID NOT RESPOND TO ANY ITEMS ON THE LCL SURVEY TO THE PROPORTION OF STUDENTS WHO DID NOT RESPOND TO ANY ITEMS ON THE HCL SURVEY

Survey Type	**# of Students Who Did Not Respond to Any Items on the Survey**	**Total # of Students**	**Proportion of Students Who Did Not Respond to Any Items on the Survey**	***z***
LCL	4	274	0.015	4.25 (p = .00)
HCL	18	277	0.065	

TABLE XIV

RESULTS FROM A *Z*-TEST TO COMPARE THE PROPORTION OF ITEMS THAT STUDENTS SKIPPED ON THE LCL SURVEY TO THE PROPORTION OF ITEMS THAT STUDENTS SKIPPED ON THE HCL SURVEY

Survey Type	**# of Items Skipped**	**Total Number of Items**	**Proportion of Items Skipped**	***z***
LCL	123	6006	0.021	9.39 (p =.00)
HCL	275	6600	0.042	

Research Question 4: Will students spend more time responding to a HCL instrument than to a LCL instrument?

The time data I collected in Primary Study 2 was not normally distributed (mean = 105.52, median = 37.55, mode = 29.20; SD = 1146.19). I removed four outlier time values from the data; but even after eliminating those extreme values, the remaining data still did not meet the assumption of normality (mean = 48.20, median = 37.27, mode = 29.20; SD = 61.50). Therefore, I chose to use a Mann-Whitney U-test, the non-parametric equivalent of a Student's *t*-test, to determine whether or not students took significantly more time to respond to the HCL version of the survey than they did to respond to the LCL version of the survey.

To obtain the outcome measures, I first calculated the total number of seconds that students spent interacting with the items on the LCL survey and the total number of seconds that students spent interacting with the items on the HCL survey. I divided the total number of seconds spent interacting with a survey by the number of students who completed that survey to obtain the mean.

I compared these two means using a Mann-Whitney U-test to determine whether the average times that students spent interacting with each survey were statistically significantly different. The formula for the Mann-Whitney U-test is described below:

$$U = R_1 - \frac{n_1(n_1+1)}{2} \quad (3)$$

where R_1 is the sum of the ranks for Sample 1 and η_1 is the sample size for Sample 1.

The results from this analysis indicated that students in Primary Study 2 spent significantly more time responding to items on the HCL survey than to items on the LCL survey ($p = .00$). However, the effect size for this statistic was small ($r = .20$), 95% CI [0.18, 0.22], indicating that the magnitude of the effect was weak.

Research Question 5: Will more students use response sets when answering items on the HCL instrument than when answering items on the LCL instrument?

The outcome measures needed to answer this research question were the numbers of students who used response sets when they took the surveys. Richard M. Smith (1996) recommended using a combination of mean-square student fit statistics and point-measure correlations from Rasch rating scale analyses to detect the presence of response sets in ratings.

Smith argued that aberrant response patterns fall into five different categories. His first category, *overfitting* (muted) response patterns may occur when a student's responses display a Guttman-like pattern[2]. This category of aberrant response will have low mean-square values (less than 1.0) and high point-measure correlations. The second category of aberrant response is *limited categories*. This includes central tendency as well as other restricted uses of the response scale. For example, in my study, some students might have shown a tendency to select only the response options that appear at the upper end of the scales. Others might have shown a tendency to select only the response options that appear at the lower end of the scales. Smith suggested that low mean-square values and high point-measure correlations are diagnostic of students' use of limited response options. Smith's third category of aberrant response is *informative-noisy*. These response patterns hold information that is useful for measuring the construct, but also contain noise, or unmodeled variance. Extreme response styles and erratic response patterns would be indicative of response sets in the informative-noisy category. This category of aberrant response can be detected by mean-square values higher than 1.0 and negative point-measure

[2] The response option selections of a student who shows a Guttman-like, deterministic response pattern are overly predictable from one another. The Rasch model predicts that for each student there should be a "zone of uncertainty or unpredictability" (Bond & Fox, 2007, p. 241) in their responses to survey items. However, the response pattern of an overfitting student does not exhibit this zone of uncertainty. When WINSTEPS analyzed the students' responses to the items included on my survey instruments, it ordered the items from those that were the easiest for the students to indicate that they were satisfied with, to those that were the hardest for students to indicate that they were satisified with. An *overfitting* student was one whose pattern of responding to the items fit this model too well (i.e., there was not much difference between the easiest item for students to be satisfied with and the most difficult). While overfit does not distort the measurement process, it could be an indication that the student was using a response set when completing the survey.

correlations. Another category of aberrant response is *non-informative*, which includes response sets that do not provide any information useful for measurement. Characteristic of this category are high mean-square values (greater than 1.5) and negative point-measure correlations. The final category of aberrant response is *contradictory*. This response pattern may occur when students do not understand the survey items. Contradictory response patterns will have large mean-square values and negative point-measure correlations.

As Smith (1996) suggested, I used a combination of mean-square error values and point-measure correlations to detect response sets. I collapsed several of his response pattern categories to make two larger categories since the detection criteria are the same for some of those categories. (That is, the detection criteria are the same for *overfitting* and *limited categories* response patterns. Similarly, the detection criteria are the same for *non-informative* and *contradictory* response patterns).

Using the Rasch rating scale model (RSM) (Andrich, 1978) to analyze the students' responses to the items, I obtained student fit statistics and point-measure correlations, which I used to detect the presence of response sets in the students' ratings. The formula for the Rasch rating scale model I used was:

$$\log (P_{nik}/P_{nik-1}) = B_n - D_i - F_k \quad (4)$$

where P_{nik} is the probability that student *n* selected response option *k* for item *i*, P_{nik-1} is the probability that student *n* selected response option *k*-1 for item *i*, B_n is the level of agreement that student *n* indicated when responding to statements about his/her college experience, D_i is the difficulty of item *i*, and F_k is the step threshold or the point at which the student was equally likely to select response option *k*-1 or response option *k*.

The output from a Rasch rating scale analysis includes for each student two different mean-square fit statistics, outfit and infit. The formula for calculating the mean-square outfit statistic is:

$$(\sum_{n=1}^{N} Z_{ni}{}^{2}) / \mathrm{N} \tag{4}$$

where N is the total number of items for which a student provided a response, and $Z_{ni}{}^{2}$ is the squared standardized residual difference between the response that student n made to item i and the expected response. Outfit statistics are based on sums of these squared standardized residuals.

The formula for calculating the mean-square infit, or information-weighted, statistic is:

$$\sum_{n=1}^{N} \mathrm{W}_{ni}\, Z_{ni}{}^{2} / \mathrm{W}_{ni} \tag{5}$$

where N is the total number of items for which a student provided a response, $Z_{ni}{}^{2}$ is the squared standardized residual difference between the response that student n made to item i and the expected response, and W_{ni} is the variance of the response of student n to item i around its expectation. The infit statistic reduces the effects of large residuals between the responses that the students made and their expected responses by dividing each residual by its variance.

The formula for calculating the point-measure correlation is:

$$r_{pmi} = \frac{\sum_{n=1}^{N}\left(X_{ni} - \bar{X}_i\right)\left(B_n - \bar{B}\right)}{\sqrt{\sum_{n=1}^{N}\left(X_{ni} - \bar{X}_i\right)^2 \sum_{n=1}^{N}\left(B_n - \bar{B}\right)^2}} \tag{6}$$

where N is the number of students, X_{ni} is the response that student n made to item i, $\bar{X}_i$ is the average of the students' responses to item i, B_n is the level of agreement that student n indicated when responding to statements regarding his/her college experience, and $\bar{B}$ is the

average level of agreement based on the responses of all students to statements regarding their college experiences.

I used the WINSTEPS computer program (Linacre, 2016) to analyze the students' responses to all the items on the LCL survey, and then I analyzed the students' responses to all the items on the HCL survey. First, I looked at the output from the principal component analysis (PCA) for each of the instruments and verified that both instruments I used met the Rasch model's assumption of unidimensionality (i.e. that each instrument measured a single underlying construct).

Next, I examined the output from these two analyses to look for aberrant response patterns, which might have indicated the use of response sets. I identified the students who showed aberrant response patterns and then classified each of those students' response patterns as fitting into one of three categories: (a) *overfitting* or *limited categories* response patterns, (b) *informative-noisy* response patterns, or (c) *non-informative* or *contradictory* response patterns.

Overall, the total number of students who displayed any pattern indicative of the use of a response set were similar for those taking the LCL survey (34 students) and those taking the HCL survey (26 students). However, as shown in Table XV, the types of response patterns that the two groups of students exhibited differed. (See Appendix D, Tables XX-XXV for more specific information.)

I used a *z*-test to compare the proportions of students taking the two surveys who had response patterns that would indicate the use of response sets. I then conducted a series of *z*-tests to compare the proportions of students taking the two surveys who exhibited these types of response patterns: (a) either *overfitting* or *limited categories* response patterns, (b) *informative-noisy* response patterns, and (c) either *non-informative* or *contradictory* response patterns.

TABLE XV

A COMPARISON OF THE NUMBERS OF RESPONSE PATTERNS THAT INDICATED THAT THOSE STUDENTS RESPONDED TO THE ITEMS USING VARIOUS TYPES OF RESPONSE SETS

Survey Type	**# (Proportion) of Response Patterns Characterized as *Overfitting* or *Limited Categories***	**# (Proportion) of Response Patterns Characterized as *Informative-Noisy***	**# (Proportion) of Response Patterns Characterized as *Non-informative* or *Contradictory***
LCL	29 (.104)	4 (.014)	1 (.0036)
HCL	10 (.036)	10 (.036)	6 (.021)

As shown in Table XVI, there was no significant difference between the proportions of students who took the HCL survey and those who took the LCL survey who exhibited response patterns indicating the use of response sets. A significantly higher proportion of students who took the LCL survey had response patterns characterized as either *overfitting* or *limited category use* (see Table XVII). By contrast, a significantly higher proportion of students who took the HCL survey had response patterns characterized as *informative-noisy* (see Table XVIII). Similarly, a significantly higher proportion of students who took the HCL survey had response patterns characterized as either *non-informative* or *contradictory* (see Table XIX). However, all of the effect sizes for these *z* statistics were small, indicating that the magnitudes of these effects were weak.

TABLE XVI

RESULTS FROM A *Z*-TEST TO COMPARE THE PROPORTIONS OF STUDENTS WITH RESPONSE PATTERNS THAT WOULD INDICATE THE USE OF RESPONSE SETS

Survey Type	**# of Students with Response Patterns That Would Indicate the Use of Response Sets**	**# of Students Responding to the Survey**	**Proportion**	***z***	**Effect Size**	**95% Confidence Interval**	
						Lower Bound	Upper Bound
LCL	34	274	0.124	-0.82 (p = .42)	r = -.36	-0.27	-0.45
HCL	26	277	0.108				

TABLE XVII

RESULTS FROM A *Z*-TEST TO COMPARE THE PROPORTIONS OF STUDENTS WITH EITHER *OVERFITTING* OR *LIMITED CATEGORIES* RESPONSE PATTERNS

Survey Type	**# of Students with Either *Overfitting* or *Limited Categories* Response Patterns**	**# of Students Responding to the Survey**	**Proportion**	***z***	**Effect Size**	**95% Confidence Interval**	
LCL	29	274	0.1060	-4.50 (p = .00)	r = -.19	Lower Bound	Upper Bound
HCL	10	277	0.0361			-0.10	-0.28

TABLE XVIII

RESULTS FROM A *Z*-TEST TO COMPARE THE PROPORTIONS OF STUDENTS WITH *INFORMATIVE-NOISY* RESPONSE PATTERNS

Survey Type	**# of Students with *Informative-Noisy* Response Patterns**	**# of Students Responding to the Survey**	**Proportion**	***z***	**Effect Size**	**95% Confidence Interval**	
LCL	4	274	0.0146	2.26 (p = .02)	r = .10	Lower Bound	Upper Bound
HCL	10	277	0.0361			0.011	0.18

TABLE XIX

RESULTS FROM A *Z*-TEST TO COMPARE THE PROPORTIONS OF STUDENTS WITH EITHER *NON-INFORMATIVE* OR *CONTRADICTORY* RESPONSE PATTERNS

Survey Type	**# of Students with Either *Non-informative* or *Contradictory* Response Patterns**	**# of Students Responding to the Survey**	**Proportion**	***z***	**Effect Size**	**95% Confidence Interval**	
LCL	1	274	0.0037	3.58 (p =.00)	r = .15	Lower Bound	Upper Bound
HCL	6	277	0.0330			0.07	0.24

V. DISCUSSION

I will open this chapter by addressing my initial research questions. I will then discuss the significance of this research, as well as its limitations. Finally, I will conclude this chapter by considering directions for future research.

A. Addressing the Research Questions

I posed five research questions to focus my research. In this section of the chapter, I will present each question, summarize the results from my studies that pertain to each one, and then, when applicable, compare my results to the results that other researchers have reported.

Research Question 1: Will students report that it takes more mental effort to respond to high cognitive load (HCL) items than to low cognitive load (LCL) items?

Overall, the students who participated in Primary Study 1 reported that the HCL items required more mental effort to respond to than the LCL items. They selected 13 of the 18 HCL versions of the items at rates that were statistically significantly higher than if they had been responding in a random manner. For 17 of the 18 items, the magnitude of each of those effects was strong, while for the remaining item, the magnitude of the effect was moderate/intermediate.

After I analyzed students' responses to individual items, I carried out a second set of analyses to examine their responses to the pair of items on the instrument that tested each recommendation. (For example, I compared the proportion of students who selected the HCL versions of Item 1 and Item 2 (which tested the first recommendation) to the proportion expected if students were to respond randomly when making their selections for those two items. I repeated these analyses for all nine pairs of items.) For each item pair, the *z* statistic was statistically significant, and the magnitude of the effect was strong.

The results from these analyses provide strong support for the inclusion of all nine of the tested recommendations under Step 1 of my proposed theoretical model (see pp. 44-45). Students perceived that the LCL versions of the items (which followed those recommendations) required less mental effort to respond to than the HCL versions of those items (which did not follow those recommendations).

While no other researchers have as yet conducted studies in which they have asked survey respondents to compare the amounts of mental effort required to respond to HCL and LCL versions of the same items, the results from my study support the views of Paas and Van Merrienboer (1994a, 1994b) who argued that the cognitive load of a task matters. My study's results suggest that survey designers can reduce the cognitive load of a survey by focusing on the task characteristics of the items and the instruments.

Research Question 2: Will students spend more time responding to a HCL instrument than to a LCL instrument?

Students participating in Primary Study 2 spent significantly more time responding to items on the HCL instrument than they spent responding to items on the LCL instrument (p = .00). However, the magnitude of the effect was weak.

This finding suggests that the HCL instrument that I created may indeed have had a higher cognitive load than the LCL instrument. That higher cognitive load may represent what Paas, Tuovinen, Tabbers, and Van Gerven (2003, p. 65) have termed *extraneous* or *ineffective load*. In the context of educational task design, they theorized that *extraneous cognitive load* results primarily from badly designed instruction, making it more difficult and time consuming for students to learn to perform a task. Translating that concept into the context of survey design, a poorly designed instrument may result in extraneous cognitive load that will make the

instrument more difficult and time consuming for respondents to complete, which, according to Krosnick (1991), may result in *satisficing* (i.e., expending less than optimal cognitive effort to respond to survey items by taking a shortcut in their cognitive processes).

Research Question 3: Will more students complete the LCL instrument than the HCL instrument?

Although there were relatively few students in Primary Study 2 who did not respond to any of the items on the instruments, the proportion of students who did not respond to any of the items on the HCL instrument was statistically significantly higher ($p = .00$) than the proportion of students who did not respond to any of the items on the LCL instrument. However, the magnitude of the effect was weak.

Krosnick (1991) and Krosnick et al. (2002) theorized that people who find a task too cognitively demanding are more likely to *satisfice,* which can take many forms. The strongest form is skipping items all together. The finding that more students taking the HCL instrument chose not to respond to any of the items suggests that there may have been cognitive features of that survey that made satisficing an attractive option. It appears that they chose the strongest form of satisfying available to students in my study who wanted to be eligible for the raffle to win an IPad Mini--providing only their demographic information, and then choosing not to respond to any of the survey items, perhaps because the task seemed too cognitively demanding.

Research Question 4: Will students respond to more items on the LCL instrument than to items on the HCL instrument?

The proportion of items that students skipped on the HCL instrument was statistically significantly higher than the proportion of items that students skipped on the LCL instrument

(p = .00), but the magnitude of the effect was weak. This finding also seems in line with the satisficing explanation that Krosnick et al. (2002) offered. Students taking the HCL instrument may have felt that more of the items on that survey exceeded the cognitive demand they were willing to exert than did students taking the LCL instrument, resulting in students' decisions to satisfice by skipping items more frequently on the HCL instrument.

Research Question 5: Will more students use response sets when answering items on the HCL instrument than when answering items on the LCL instrument?

Overall, the total number of students participating in Primary Study 2 who displayed any pattern indicative of the use of a response set was similar for those answering items on the LCL instrument and those answering items on the HCL instrument. However, the types of aberrant response patterns that students exhibited differed for those two groups. A significantly higher proportion of students who answered items on the LCL instrument had response patterns characterized as either *overfitting* or *limited category use*. By contrast, a significantly higher proportion of students who answered items on the HCL instrument had response patterns characterized as *informative-noisy*. Similarly, a significantly higher proportion of students who answered items on the HCL instrument had response patterns characterized as either *non-informative* or *contradictory*. However, the magnitude of each of these effects was weak.

These results suggest that some of the students may have exhibited evidence of *acquiescence bias* (Cannell et al., 1981; Krosnick, 1991). That is, when responding to a series of statements using an agreement scale, these students may have tended to agree with most (or all) of those statements, perhaps because, in so doing, they felt that they were making socially desirable choices. Their response patterns would *overfit* the measurement model because they

were only using the response options that appear at the upper end of the agreement scale.[3] The response patterns of about 11% of the students taking the LCL version of the survey displayed evidence of overfit, but only about 4% of the students taking the HCL version had overfitting response patterns. While these findings suggest that students taking the LCL instrument may have been more prone to acquiescence bias, it is not clear why that should have been the case.

However, the other types of response sets that students taking the HCL version of the survey more frequently displayed (i.e., *informative-noisy*, *non-informative* or *contradictory*) are potentially more problematic since they signal that those students' response patterns were rather noisy and erratic, lacking predictability. Moreover, the amount of useful information that a survey designer can obtain from those students' responses to the items can often be very limited. My results suggest that students taking HCL instruments may be somewhat more likely than students taking LCL instruments to exhibit evidence of using these more problematic response sets.

A. Limitations of the Research

These studies of cognitive load and its role in survey design have produced some useful, practical information to guide survey construction. However, the studies had several limitations. It is important to take these limitations into account when interpreting the results from the studies, determining the merits of the conclusions drawn, and planning future research that would seek to test the efficacy of the proposed theoretical model for reducing cognitive load in survey instruments.

First, for this set of studies, I only tested guidelines and recommendations that pertain to Step 1 (Interpretation) of my theoretical model. In my model, I included guidelines and

[3] Overfit might also occur if students were only using the response options that appear at the lower end of the agreement scale.

recommendations related to each of the four cognitive steps that Tourangeau (1984) proposed to explain the role of cognition in survey response. While some of the guidelines and recommendations appear under multiple steps in my model, others do not. Therefore, it is important to acknowledge that my studies do not test all the guidelines or recommendations in my proposed model, but rather only a subset of them.

When designing my studies, I decided to work with a population that was familiar to me: students. As an assessment administrator in higher education, I had ready access to this population. I chose to create online survey instruments to measure a construct that I hoped would be of interest to students (i.e., their level of satisfaction with their college experiences). I created a series of statements for students to respond to using a 4-point agreement scale and then used their survey responses to test some of the guidelines and recommendations in my theoretical model. However, the results from my studies may not be generalizable to other populations, types of surveys, survey response formats, and/or survey delivery modes. For example, the demographic characteristics of my samples were likely different from those of a more general population in terms of educational level and socioeconomic status. To what extent does that limit the generalizability of the findings from my studies? It would be useful to test the robustness of the guidelines and recommendations in my theoretical model using other samples besides college students, as well as other types of surveys, survey response formats, and survey delivery modes.

All statistical analyses have inherent strengths and weaknesses and make certain assumptions about both the population and the samples studied. I had initially planned to use a Student's *t*-test to compare the amounts of time that students spent responding to the HCL and LCL versions of the instruments. However, use of this statistical test assumes that the data are sampled from a normally distributed population, and that the sample represents a random sample

from a defined population. My time data violated the assumption of normality. Therefore, I decided to use the nonparametric analog for the Student's *t*-test, the Mann-Whitney U-test, to analyze the time data. While the Mann-Whitney U-test provided more interpretable results, the time data remains problematic because I administered the survey online. Although I asked students to complete the survey in one sitting, it is possible that not every student complied with this request. Additionally, if a student took no time (or a very small amount of time) to respond to a particular item, that may indicate that the item had a high cognitive load and the student may have chosen to skip the item all together, significantly reducing the time that the student spent responding to the survey instrument.

B. Significance of the Research

Despite the limitations of my studies, I believe this work contributes in meaningful ways to the larger body of research on survey design. My studies take as their starting point Tourangeau's (1984) model of the cognitive processes involved in survey responding. I used his model as the basis for the theoretical model that I proposed for reducing cognitive load in survey instruments. I also included in my model many of the guidelines and recommendations to reduce cognitive effort in survey responding that Tourangeau and other survey researchers such as Cannell and his colleagues (1981) have suggested. In this way, my work builds on and extends the work of researchers who have conducted studies to examine the role of cognition in responding to surveys.

My studies also build on the work of educational psychologists who have proposed various theories of cognitive load and have carried out research using different methods to try to measure it, primarily in education-related settings. While most of the researchers working in this area have focused on the educational implications of cognitive load, my research examines the

role that cognitive load plays in survey design, demonstrating the practical utility of this concept in this particular context. My hope is that, through this research, practitioners who are constructing surveys may begin to gain an understanding of cognitive load and how it might impact survey design.

Perhaps most importantly, my studies serve to bridge these two disparate bodies of research: the survey design literature, and research on cognitive load. Researchers working in these two areas have made invaluable contributions to our understanding of cognitive load and survey design; but the two groups of researchers have worked in parallel, making little or no attempt over the years to collaborate, or even to borrow ideas and learn from one another. My studies represent a first attempt to design research that draws upon key findings from both literatures.

The theoretical model that I proposed represents a first attempt to identify survey design guidelines and recommendations that relate to each of the four steps in Tourangeau's model of the cognitive processes involved in survey responding. Currently, there is no practical, accessible, easy-to-use model that practitioners can use to design survey instruments. Most practitioners who are creating surveys certainly want to create instruments that will produce valid and reliable results. However, when looking for guidance regarding best practices in survey design, they are faced with a large, confusing, and often inaccessible body of literature on survey design that provides myriad guidelines and recommendations, some of which have little or no research to support them. To make matters worse, in some instances when researchers have carried out research to test certain guidelines and recommendations, they have reported conflicting results. The model I have proposed could prove very useful to practitioners, especially if, over time, researchers are able to provide convincing evidence that employing this

model leads to reduced cognitive load and improved survey outcomes. While I was only able to test some of the guidelines and recommendations in my model using an online survey administered to college students, the initial results seem promising; but, clearly, researchers will need to conduct additional studies in order to test the efficacy of the full model when applied in varied contexts.

C. Directions for Future Research

This project represents a first attempt to plan and design studies that hopefully will be part of a coordinated, long-term research agenda aimed at exploring the utility of my theoretical model for reducing cognitive load, a model that seeks to identify a series of best practices for survey design. In the future, researchers could build upon this project, planning and carrying out additional studies aimed at refining the model.

My research focused primarily on gathering and analyzing instrument-level data. However, conducting item-by-item analyses may provide additional insights into the ways in which cognitive load affects responses to survey items. For example, further investigation into different items that follow the same guidelines and/or recommendations may help to refine those recommendations/guidelines. Also, analyses of item-level data may clarify which particular guidelines and recommendations are important for survey designers to follow if their goal is to reduce the amount of cognitive effort required to respond to items.

In the pilot study, I individually interviewed students, presenting them with two versions of each item and then asking them to think out loud as they decided which version would require more mental effort to respond to. The rationales that they gave to support their choices provided some interesting and useful insights into their perceptions of cognitive load. In the future,

researchers might consider replicating this study using larger and more representative student samples to determine whether those students would provide similar (or different) rationales to support their choices. If the initial insights I gained from my interviews with this small sample of students could be verified with larger samples, then the theoretical model that I proposed could afford much more flexibility as a survey design model than the current guidance available to practitioners (which tends to be lists of individual recommendations). Since the preliminary findings seemed to indicate that what creates greater cognitive load may depend, at least to some extent, on the specific population being studied, use of the model that I am proposing could give practitioners more flexibility to adjust individual recommendations based on the specific populations they aim to study.

In my studies, I was only able to test guidelines and recommendations that pertain to one of the four steps in the theoretical model, the interpretation step. Researchers will also need to test the guidelines and recommendations for the remaining three steps (i.e., recall, judgment, and response). While there is some overlap in the guidelines and recommendations included in those steps, there are also some guidelines and recommendations that are unique to certain steps that I was not able to test. It might also be informative to conduct item-by-item analyses to determine if the items created using different guidelines and recommendations differ in terms of the distributions of responses to those items. Results from these types of analyses may provide information that could help refine the model.

Researchers who are interested in studying cognitive load and its impact on survey design may also want to consider using the theoretical model to create HCL and LCL versions of surveys that measure a variety of constructs and/or that use response options other than an agreement scale. Researchers may also want to explore the utility of this model for creating

different types of surveys besides those that measure satisfaction. For example, many researchers are interested in surveys that contain fact-based items. Is it possible to use the guidelines and recommendations in the model to construct HCL and LCL versions of fact-based items? If so, will respondents show different patterns of responding to fact-based survey instruments that are designed to be parallel in content but differ in their cognitive loads? Finally, researchers might consider administering HCL and LCL versions of surveys using samples other than students and/or experimenting with different formats for delivering those surveys (i.e., paper-and-pencil) to test the utility and robustness of the guidelines and recommendations in diverse contexts.

The question of whether those who take HCL versions of surveys use response sets more frequently than those who take LCL versions of surveys is an intriguing one that is worthy of further investigation. In my study, the total numbers of students who displayed any pattern indicative of the use of a response set were similar for those taking the LCL version of the instrument and those taking the HCL version. However, the types of aberrant response patterns that those two groups of students exhibited differed. Students taking the LCL version displayed more *overfitting* and *limited categories* response patterns than students taking the HCL version. By contrast, students taking the HCL version displayed more *informative-noisy*, *non-informative*, and *contradictory* response patterns than students taking the LCL version. Are these findings unique to the particular context I studied? Would researchers using surveys designed to measure other constructs report similar or different findings? Would researchers studying other samples of survey respondents report similar or different findings? Additional studies to explore the extent to which these preliminary findings are generalizable represent an exciting opportunity for

researchers to learn more about whether differences in the cognitive load of items will result in differences in the patterns and uses of various types of response sets.

REFERENCES

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods and Research, 20*(1), 139-181. doi: 10.1177/04912419020001005

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly, 48*, 409-442. doi: 10.1086/268840

Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality*. New York, NY: Plenum Press.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43,* 561-73.

Aper, J. P., & Hinkle, D. E. (1991). State policies for assessing student outcomes: A case study with implications for state and institutional authorities. *The Journal of Higher Education*, *62*, 539-555.

Astin, A. (1990). Educational assessment and educational equity. *American Journal of Education*, *98*, 458-478. doi: 10.1002/aehe.3640220309

Bendig, A. W. (1953). The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. *Journal of Applied Psychology, 37*, 38-41. doi: 10.1037/h0057911

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research, 14*(2), 280-288.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Burke, J. C., & Modarresi, S. (2000). To keep or not to keep performance funding: Signals from stakeholders. *The Journal of Higher Education*, *71*, 432-453. doi: 10.1057/palgrave.hep.8300096

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*, 389-437.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cook, C., Heath, F., & Russel, L.T. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement, 60*, 821-836. doi:10.1177/00131640021970934

Couch, A., & Kenniston, K. (1960). Yaysayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology, 80*(2), 151-174.

Couper, M. P., Tourangeau, R., & Kenyon, K. (2004). Picture this! Exploring visual effects in web surveys. *Public Opinion Quarterly, 68*(2), 255-256.

Cox, E. P., III (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research, 17*(4), 407-422.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). The basics of crafting good questions. In *Internet, mail, and mixed-mode surveys: The tailored design method* (pp. 65-106), New York, NY: Wiley.

Elton, L. (1988). Accountability in higher education: The danger of unintended consequences. *Higher Education*, *17*, 377-390.

Ewell, P. T. (1991). To capture the ineffable: New forms of assessment in higher education. *Review of Research in Education*, *17*, 75-125.

Feick, L. F. (1989). Latent class analysis of survey questions that include don't know responses. *Public Opinion Quarterly, 53*, 525-547.

Fischer, K. (2006). Lengthy fights are expected over measures on accountability. *The Chronicle of Higher Education, 53*(2), 42.

Fowler, F. J. (1995). Some general rules for designing good survey instruments. In *Improving survey questions: Design and evaluation* (pp.78-103). *Applied Research Methods Series, 13*. Thousand Oaks, CA: Sage Publications.

Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2007). Eye tracking data: New insights on response order effects and other cognitive shortcuts in survey responding. *Public Opinion Quarterly, 72*(5), 892-913.

Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research, 29*(2), 176-188.

Greenleaf, E. A. (1992b). Measuring extreme response style. *Public Opinion Quarterly, 56*, 328-351.

Hodson, P., & Thomas, H. (2003). Quality assurance in higher education: Fit for the new millennium of simply year 2000 compliant? *Higher Education*, *45*, 375-387.

Holm, S. (1979). A simple sequential rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65-70.

Jahoda, M., Deutsch, M., & Cook, S. (1951). *Research methods in social relations* (Vol. 2). Champaign, IL: University of Illinois Press.

Jenkins, C. R., & Dillman, D. A. (1997). Toward a theory of self-administered questionnaire design. In L. E. Lyberg, P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 165-198). Hoboken, NJ: John Wiley & Sons, Inc.

Jenkins, G. D., & Taber, T. A. (1977). Monte Carlo study of factors affecting three indices of composite scale reliability. *Journal of Applied Psychology, 62*(4), 392-398.

Jobe, J. B., & Mingay, D. J. (1991). Cognition and survey measurement: History and overview. *Applied Cognitive Psychology, 5*(3), 175-192.

Johnson, T., O'Rourke, D., Chavez, N., Sudman, S., Warnecke, R., & Lacey, L. (1997). In L. E. Lyberg, P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 165-198). Hoboken, NJ: John Wiley & Sons, Inc.

Kearns, K. P. (1998). Institutional accountability in higher education: A strategic approach. *Public Productivity and Management Review*, *22*, 140-156.

King Alexander, F. (2000). The changing face of accountability: Monitoring and assessing institutional performance. *The Journal of Higher Education*, *71*, 411-431.

Kingkade, T. (2012). Scott Walker calls for shift in higher education funding tying dollars to completion rates. *Huffington Post.* Retrieved from http://www.huffingtonpost.com/2012/11/19/scott-walker-wisconsin-higher-education-funding_n_2159907.html

Klopfer, F. J., & Madden, T. M. (1980). The middlemost choice on attitude items: Ambivalence, neutrality, or uncertainty? *Personality and Social Psychology Bulletin, 6*, 97-101. doi: 10.1177/014616728061014

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of scales. *Educational and Psychological Measurement, 25*, 987-995. doi:10.1177/001316446502500404

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*, 213-236. doi: 10.1002/acp.2350050305

Krosnick, J. (1999). Survey research. *Annual Review of Psychology, 50,* 537-567.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., . . . Conaway, M. (2002). The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *The Public Opinion Quarterly, 66*(3), 371-403. DOI: 10.1.1.141.7834

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities.* Retrieved from http://learningoutcomesassessment.org/documents/NILOAStateStudy_000.pdf

Lehmann, D. R., & Hulbert, J. (1972). Are three-point scales always good enough? *Journal of Marketing Research*, *9*(4), 444-446.

Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes*. Bibergau, Germany: Psychometrica. Retrieved from http://www.psychometrica.de/effect_size.html. doi: 10.13140/RG.2.1.3478.4245

Leveille, D. E. (2005). *An emerging view on accountability in American higher education*. Retrieved from the University of California, Berkeley, Center for Studies in Higher Education Web site: http://cshe.berkeley.edu/publications/publications.php?id=54

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M. (2016). WINSTEPS (Version 3.92.1)[Computer software]. Minnetonka, MN: SWREG DR MyCommerce, Inc.

Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology, 60*(1), 10-13. doi: 10.1037/h0076268

MetaMetrics. (2015). *Lexile-to-grade correspondence*. Retrieved from: https://www.lexile.com/about-lexile/grade-equivalent/grade-equivalent-chart/

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

Murphy, G., & Likert, R. (1938). *Public opinion and the individual.* New York, NY: Harper and Brothers.

Neave, G. N. (1980). Accountability and control. *European Journal of Education*, *15*, 49-60.

New Leadership Alliance for Student Learning and Accountability. (2012). *Committing to quality: Guidelines for assessment and accountability in higher education.* Washington, DC: Author.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.

Oksenberg, L., & Cannell, C. F. (1977). Some factors underlying the validity of response in self report. *International Statistical Bulletin,* 48, 324-346.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill 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 developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P.W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1994b). Validity of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133.

Peter, J. P. (1979). Reliability: A review of psychometric basics and recent marketing practices. *Journal of Marketing Research, 16*, 6-17.

Principles for effective assessment of student achievement. (2013, July 19). Retrieved from http://www.ncahlc.org

Qualtrics (Version 61472 of the Qualtrics Research Suite). (2013). Provo, UT: Qualtrics.

Schuman, H., & Presser, S. (1977). Question working as an independent variable in survey analysis. *Sociological Methods and Research, 6*, 151-170.

Schwarz, N., Knauper, B., Hippler, H. J., Noelle-Neumann, E., & Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *The Public Opinion Quarterly, 4*, 570-582.

Schwarz, N., Strack, F., Hippler, H. J., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5*, 193-212.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychology Review, 63*(2), 129-138.

Smith, R. M. (1996). Polytomous mean-square fit statistics. *Rasch Measurement Transactions*, *10*(3), 516-517.

Smith, T. W., & Peterson, B. L. (1985, August). *The impact of number of response categories on inter-item associations: Experimental and simulated results*. Paper presented at the American Sociological Association Meeting, Washington, D.C.

Stratford, M. (2014, January 30). Call to action on accreditation. *Inside Higher Education*. Retrieved from http://www.insidehighered.com/news/2014/01/30/accreditors-weigh-growing-political-debate-dc-over-their-field

Sudman, S., & Bradburn, N. (1974). *Response effects in surveys: A review and synthesis.* Chicago, IL: Aldine.

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

Symonds, P. (1924). On the loss of reliability in ratings due to coarseness in scale. *Journal of Experimental Psychology, 7*, 456-461.

Tourangeau, R. (1984). Cognitive science and survey methods: A cognitive perspective. In T. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines.* Washington, D.C.: National Academy Press.

Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly, 71*(1), 91-112.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*(3), 299-314.

U.S. Department of Education. (2008). *Higher Education Opportunity Act - 2008*. Retrieved from http://www2.ed.gov/policy/highered/leg/hea08/

U.S. Department of Education, National Center for Education Statistics. (2014). *The Condition of Education 2014* (NCES 2014-083). Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=40

Van Valey, T. L. (2001). Recent changes in higher education and their ethical implications. *Teaching Sociology*, *29*, 1-8.

Visser, P. S., Krosnick, J. A., Marquette, J, & Curtin, M. (1996). Mail surveys for election forecasting? An evaluation of the Columbus Dispatch Poll. *Public Opinion Quarterly, 60,* 181-227.

Wanke, M., & Schwarz, N. (1997). Reducing question order effects: The operation of buffer items. In L. E. Lyberg, P. Biemer, M. Collins, E. D. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 115-140). Hoboken, NJ: John Wiley & Sons, Inc.

Wejters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *Vlerick Leuven Gent Working Paper Series*, Ghent, Belgium: Vlerick Leuven Gent Management School.

What you need to know about reauthorization. (2013, September 19). *Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/What-You-Need-to-Know-About/141697

White House. (2013). *College scorecard*. Retrieved from http://www.whitehouse.gov/issues/education/higher-education/college-score-card

APPENDICES

APPENDIX A

Pilot Study Instrument

Please click on your participant number below
1
2
3
4
5
6
7
8
9
10

Directions: You will see two versions of a series of survey items. The versions may vary in terms of their wording, sentence structure, and/or design. Please click on the version of the survey item that requires more mental effort (or thinking) to answer. Mental effort is how much you have to think about the item to give a response. For example, if an item requires more mental effort, you might have to read it twice. If you do not feel comfortable answering any item, you may skip it.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- o I am satisfied with the availability of career planning services at my institution.
- o I am satisfied with the career planning services I have received at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- o I am satisfied with the different options to register for classes at my institution.
- o I am satisfied with the process to register for classes at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- o I am not satisfied with the types of assignments required in my classes.
- o I am satisfied with the types of assignments required in my classes.

APPENDIX A (continued)

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the diversity of perspectives that I have been exposed to at my institution.
- I am not satisfied with the diversity of perspectives that I have been exposed to at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- Thinking about all the ways I use the internet, including web browsing, email, and social media, I am satisfied with the internet connectivity on campus.
- I am satisfied with the internet connectivity on campus.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the opportunities for community service at my institution.
- Thinking about all the various opportunities that my institution makes available for me to participate in community service, I am satisfied with the opportunities for community service at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- There is a strong sense of school spirit at my institution.
- *There is a strong sense of school spirit at my institution.*

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- *My instructors show concern for how much I am learning in my classes.*
- My instructors show concern for how much I am learning in my classes.

APPENDIX A (continued)

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with my advisor's aptitude in addressing my needs.
- I am satisfied with my advisor's ability to address my needs.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My institution communicates effectively with students about safety issues.
- My institution effectively transmits to students information regarding safety issues.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the academic challenge of the courses at my institution.
- I am satisfied with the academic rigor of the courses at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My instructors provide me with feedback that helps me improve.
- My instructors provide me with formative feedback.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I can easily navigate my institution's LMS.
- I can easily navigate my institution's learning management system (for example, Blackboard, Canvas, Desire2Learn, Sakai, Moodle).

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My institution has sufficient information available about applying for federal financial aid.
- My institution has sufficient information available about completing the FAFSA.

APPENDIX A (continued)

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My institution provides students with sufficient information about what to do in case of an emergency.
- My institution provides students with sufficient information about its emergency preparedness plan.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the assistance that my institution provides to support my computer usage.
- I am satisfied with the infrastructure support that my institution provides to support my computer usage.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the access I have to academic journals at my institution.
- Considering the access to academic journals I have at my institution, I am satisfied.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My instructors have good knowledge about the content they teach.
- Based on my experiences with instructors with whom I have taken classes, instructors at my institution have good knowledge about the content they teach.

APPENDIX B

Primary Study 1 Instrument

Below, you will see a link to a document called an "Informed Consent" document. This document provides valuable information about this research, including why are you being asked to participate, the purpose and procedures, potential risks and benefits, other options to participation, whom to contact if you have questions or concerns, and your ability to withdraw from the research. Please read this document carefully and print a copy for yourself.

[Link to Informed Consent]

After you have read the Informed Consent document, click on one of the options below to indicate whether or not you consent to participate in this research.

- o I consent to participating in this study.
- o I DO NOT consent to participating in this study.

Please indicate your gender by clicking on one of the options below.

- o Male
- o Female
- o Other (textbox)

Please type the year in which you were born*
*Please note you must be at least 18 years old to participate in this survey.
(textbox)

Please select the option or options that best describe your race/ethnicity (click on all that apply).

- o American Indian or Alaska Native
- o Asian
- o Black or African American
- o Hispanic or Latino
- o Native Hawaiian or Pacific Islander
- o White

Please click on the option that best describes the type of higher education institution you attend.

- o 2 year/community college
- o 4 year public college or university
- o 4 year private college or university
- o Other (textbox)

APPENDIX B (continued)

What is the highest level of education you have completed? Please click on one of the options below.

- o High School Diploma or GED
- o Associate's Degree or Equivalent
- o Bachelor's Degree or Equivalent
- o Master's Degree or Equivalent
- o Doctoral Degree or Equivalent Professional Degree
- o Other (textbox)

Please indicate the type of education you are currently seeking (click all that apply).

- o Associate's Degree
- o Bachelor's Degree
- o Master's Degree
- o Doctoral Degree or Equivalent Professional Degree
- o Other (textbox)

Directions: You will see two versions of a series of survey items. The versions may vary in terms of their wording, sentence structure, and/or design. Please click on the version of the item that requires **more** mental effort (or thinking) to answer. Mental effort is how much you have to think about the item to give a response. For example, if an item requires more mental effort, you might have to read it twice. If you do not feel comfortable answering any item, you may skip it.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- o I am satisfied with the availability of tutoring services at my institution.
- o I am satisfied with the tutoring services I have received at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- o I am satisfied with the availability of resources to apply for on-campus jobs.
- o I am satisfied with the resources I have used to apply for on-campus jobs.

APPENDIX B (continued)

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am not satisfied with the types of assignments required in my classes.
- I am satisfied with the types of assignments required in my classes.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the diversity of perspectives that I have been exposed to at my institution.
- I am not satisfied with the diversity of perspectives that I have been exposed to at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the internet connectivity on campus.
- Thinking about all the ways I use the internet, I am satisfied with the internet connectivity on campus.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the opportunities for community service at my institution.
- Thinking about all the various opportunities that my institution makes available for me to participate in community service, I am satisfied with the opportunities for community service at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- *There is a strong sense of school spirit at my institution.*
- There is a strong sense of school spirit at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My instructors show concern for how much I am learning in my classes.
- *My instructors show concern for how much I am learning in my classes.*

APPENDIX B (continued)

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My institution's campus bookstore stocks an adequate assortment of textbooks.
- My institution's campus bookstore stocks an adequate mélange of textbooks.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My institution communicates effectively with students about safety issues.
- My institution effectively transmits to students information regarding safety issues.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the academic challenge of the courses at my institution.
- I am satisfied with the academic rigor of the courses at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My instructors provide me with feedback that helps me improve.
- My instructors provide me with formative feedback.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I can easily navigate my institution's LMS.
- I can easily navigate my institution's learning management system (for example, Blackboard, Canvas, Desire2Learn, Sakai, Moodle).

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the cost of attendance at my institution.
- I am satisfied with the COA at my institution.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My institution provides students with sufficient information about what to do in case of an emergency.
- My institution provides students with sufficient information about its emergency preparedness plan.

APPENDIX B (continued)

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the assistance that my institution provides to support my computer usage.
- I am satisfied with the infrastructure support that my institution provides to support my computer usage.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- I am satisfied with the access I have to academic journals at my institution.
- Considering the access to academic journals I have at my institution, I am satisfied.

Please click on the version of the survey item that requires **more** mental effort to provide a response. The response options for both items are: Strongly Agree; Agree; Disagree; Strongly Disagree; Not Applicable.

- My instructors have good knowledge about the content they teach.
- Based on my experiences with instructors with whom I have taken classes, instructors at my institution have good knowledge about the content they teach.

If you are interested in being entered in the drawing to win one of three $50 Amazon.com gift cards, please click on this link (or copy and paste it into your internet browser):

*Note you are being directed to another survey form to enter your information so your contact information will not be associated with your responses to this survey.

APPENDIX C

Primary Study 2 Instruments

HCL Instrument

Below, you will see a link to a document called an "Informed Consent" document. This document provides valuable information about this research, including why are you being asked to participate, the purpose and procedures, potential risks and benefits, other options to participation, whom to contact if you have questions or concerns, and your ability to withdraw from the research. Please read this document carefully and print a copy for yourself.

[Link to Informed Consent]

After you have read the Informed Consent document, click on one of the options below to indicate whether or not you consent to participate in this research.

- o I consent to participating in this study.
- o I DO NOT consent to participating in this study.

Please indicate your gender by clicking on one of the options below.

- o Male
- o Female
- o Other (textbox)

Please type the year in which you were born*
*Please note you must be at least 18 years old to participate in this survey.
(textbox)

Please select the option or options that best describe your race/ethnicity (click on all that apply).

- o American Indian or Alaska Native
- o Asian
- o Black or African American
- o Hispanic or Latino
- o Native Hawaiian or Pacific Islander
- o White

Please click on the option that best describes the type of higher education institution you attend.

- o 2 year/community college
- o 4 year public college or university
- o 4 year private college or university
- o Other (textbox)

APPENDIX C (continued)

What is the highest level of education you have completed? Please click on one of the options below.

- o High School Diploma or GED
- o Associate's Degree or Equivalent
- o Bachelor's Degree or Equivalent
- o Master's Degree or Equivalent
- o Doctoral Degree or Equivalent Professional Degree
- o Other (textbox)

Please indicate the type of education you are currently seeking (click all that apply).

- o Associate's Degree
- o Bachelor's Degree
- o Master's Degree
- o Doctoral Degree or Equivalent Professional Degree
- o Other (textbox)

Directions: Please read each statement and then click on the option that indicates your level of agreement with that statement. This is a brief survey; please complete it in one sitting if possible. If you do not feel comfortable responding to a particular statement, you may skip it.

I am satisfied with the tutoring services I have received at my institution.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the resources I have used to apply for on-campus jobs.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am not satisfied with the types of assignments required in my classes.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

I am not satisfied with the diversity of perspectives that I have been exposed to at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

Thinking about all the ways I use the internet, I am satisfied with the internet connectivity on campus.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

Thinking about all the various opportunities that my institution makes available for me to participate in community service, I am satisfied with the opportunities for community service at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

There is a strong sense of school spirit at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My instructors show concern for how much I am learning in my classes.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My institution's campus bookstore stocks an adequate mélange of textbooks.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

My institution effectively transmits to students information regarding safety issues.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the academic rigor of the courses at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My instructors provide me with formative feedback.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I can easily navigate my institution's LMS.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the COA at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My institution provides students with sufficient information about its emergency preparedness plan.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

I am satisfied with the infrastructure support that my institution provides to support my computer usage.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

Considering the access to academic journals I have at my institution, I am satisfied.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

Based on my experiences with instructors with whom I have taken classes, instructors at my institution have good knowledge about the content they teach.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the variety and availability of elective (i.e., non-required) courses at my institution.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the types of activities available for me to participate in at my institution and the times that those activities are scheduled.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

My degree will prepare me well for the career path that I would like to pursue. (respond using drop-down box)

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

The other students in my courses help me to learn course material better. (respond using drop-down box)

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

Thank you for participating in this research. When you consented to take part in this study, you were told that the purpose of this research was to determine how satisfied students are with various experiences that the higher education institutions they attend have provided. However, this is not the true purpose. Please click on the link below to see information about the true purpose of this study, the reason this purpose was not revealed before you provided consent, and your option to withdraw consent if this changes your decision about consenting to participate in this research.

[link to High Cognitive Load Explanation of True Purpose of Research]

Given what you now know about the true purpose of this research, please click on the appropriate option below to indicate if you would like to maintain your consent to participate in the research or withdraw your consent and have your responses to the survey removed from the research database.

- o I would like to maintain my consent to participate in this research
- o I would like to withdraw my consent to participate in this research

Thank you for participating in this survey. If you are interested in being entered in the drawing to win an IPad Mini, please click on this link (or copy and paste it into your internet browser):

*Note you are being directed to another survey form to enter your information so your contact information will not be associated with your responses to this survey.

APPENDIX C (continued)

LCL Instrument

Below, you will see a link to a document called an "Informed Consent" document. This document provides valuable information about this research, including why are you being asked to participate, the purpose and procedures, potential risks and benefits, other options to participation, who to contact if you have questions or concerns, and your ability to withdraw from the research. Please read this document carefully and print a copy for yourself.

[Link to Informed Consent]

After you have read the Informed Consent document, click on one of the options below to indicate whether or not you consent to participate in this research.

- o I consent to participating in this study.
- o I DO NOT consent to participating in this study.

Please indicate your gender by clicking on one of the options below.

- o Male
- o Female
- o Other (textbox)

Please type the year in which you were born*
*Please note you must be at least 18 years old to participate in this survey.
(textbox)

Please select the option or options that best describe your race/ethnicity (click on all that apply).

- o American Indian or Alaska Native
- o Asian
- o Black or African American
- o Hispanic or Latino
- o Native Hawaiian or Pacific Islander
- o White

Please click on the option that best describes the type of higher education institution you attend.

- o 2 year/community college
- o 4 year public college or university
- o 4 year private college or university
- o Other (textbox)

APPENDIX C (continued)

What is the highest level of education you have completed? Please click on one of the options below.

- o High School Diploma or GED
- o Associate's Degree or Equivalent
- o Bachelor's Degree or Equivalent
- o Master's Degree or Equivalent
- o Doctoral Degree or Equivalent Professional Degree
- o Other (textbox)

Please indicate the type of education you are currently seeking (click all that apply).

- o Associate's Degree
- o Bachelor's Degree
- o Master's Degree
- o Doctoral Degree or Equivalent Professional Degree
- o Other (textbox)

Directions: Please read each statement and then click on the option that indicates your level of agreement with that statement. This is a brief survey; please complete it in one sitting if possible. If you do not feel comfortable responding to a particular statement, you may skip it.

I am satisfied with the availability of tutoring services at my institution.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the availability of resources to apply for on-campus jobs.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the types of assignments required in my classes.

- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

I am satisfied with the diversity of perspectives that I have been exposed to at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the internet connectivity on campus.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the opportunities for community service at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

There is a strong sense of school spirit at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My instructors show concern for how much I am learning in my classes.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My institution's campus bookstore stocks an adequate mélange of textbooks.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

My institution communicates effectively with students about safety issues.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the academic challenge of the courses at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My instructors provide me with feedback that helps me improve.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I can easily navigate my institution's learning management system (for example, Blackboard, Canvas, Desire2Learn, Sakai, Moodle).
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the cost of attendance at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My institution provides students with sufficient information about what to do in case of an emergency.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

I am satisfied with the assistance that my institution provides to support my computer usage.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the access I have to academic journals at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My instructors have good knowledge about the content they teach.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with the variety and availability of elective (i.e., non-required) courses at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

I am satisfied with:

the types of activities available for me to participate in at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

the times that those activities are scheduled.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

APPENDIX C (continued)

I am satisfied with:

the variety of elective (i.e., non-required) courses at my institution.
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable
- o

the availability of elective (i.e., non-required) courses at my institution
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

My degree will prepare me well for the career path that I would like to pursue. (respond using radio buttons)
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

The other students in my courses help me to learn course material better. (respond using radio buttons)
- o Strongly Agree
- o Agree
- o Disagree
- o Strongly Disagree
- o Not Applicable

Thank you for participating in this research. When you consented to take part in this study, you were told that the purpose of this research was to determine how satisfied students are with various experiences that the higher education institutions they attend have provided. However, this is not the true purpose. Please click on the link below to see information about the true purpose of this study, the reason this purpose was not revealed before you provided consent, and your option to withdraw consent if this changes your decision about consenting to participate in this research.

[link to Low Cognitive Load Explanation of True Purpose of Research]

Given what you now know about the true purpose of this research, please click on the appropriate option below to indicate if you would like to maintain your consent to participate in the research or withdraw your consent and have your responses to the survey removed from the research database.

- o I would like to maintain my consent to participate in this research
- o I would like to withdraw my consent to participate in this research

Thank you for participating in this survey. If you are interested in being entered in the drawing to win an IPad Mini, please click on this link (or copy and paste it into your internet browser):

*Note you are being directed to another survey form to enter your information so your contact information will not be associated with your responses to this survey.

APPENDIX D

TABLE XX

STUDENTS WHO TOOK THE LCL SURVEY WHO DISPLAYED *OVERFITTING* OR *LIMITED CATEGORIES* RESPONSE PATTERNS

Student #	**MNSQ Outfit**	**MNSQ Infit**	**Point-Measure Correlation**
15	0.14	0.15	0.61
26	0.77	0.73	0.66
27	0.46	0.45	0.69
37	0.52	0.50	0.62
39	0.25	0.24	0.69
52	0.38	0.37	0.64
66	0.64	0.67	0.62
79	0.32	0.36	0.64
98	0.62	0.65	0.62
105	0.61	0.64	0.60
106	0.81	0.78	0.62
110	0.35	0.36	0.65
113	0.68	0.68	0.63
117	0.24	0.24	0.64
130	0.58	0.57	0.65
141	0.65	0.65	0.63
151	0.28	0.28	0.73
158	0.61	0.62	0.69
173	0.74	0.64	0.62
187	0.61	0.59	0.78
192	0.33	0.32	0.60
197	0.46	0.46	0.66
202	0.64	0.67	0.66
211	0.80	0.79	0.72
218	0.49	0.48	0.63
233	0.68	0.70	0.73
236	0.72	0.78	0.65
239	0.39	0.37	0.66
257	0.78	0.82	0.65

APPENDIX D (continued)

TABLE XXI

STUDENTS WHO TOOK THE LCL SURVEY WHO DISPLAYED *INFORMATIVE-NOISY* RESPONSE PATTERNS

Student #	MNSQ Outfit	MNSQ Infit	Point-Measure Correlation
1	9.90	5.76	-0.04
3	1.08	1.12	-0.96
43	1.06	1.07	-0.03
171	1.37	1.36	-0.10

TABLE XXII

STUDENTS WHO TOOK THE LCL SURVEY WHO DISPLAYED *NON-INFORMATIVE* OR *CONTRADICTORY* RESPONSE PATTERNS

Student #	MNSQ Outfit	MNSQ Infit	Point-Measure Correlation
1	9.9	5.76	-0.04

TABLE XXIII

STUDENTS WHO TOOK THE HCL SURVEY WHO DISPLAYED *OVERFITTING* OR *LIMITED CATEGORIES* RESPONSE PATTERNS

Student #	MNSQ Outfit	MNSQ Infit	Point-Measure Correlation
49	0.73	-1.0	0.62
8	0.46	-1.7	0.63
218	0.57	-1.4	0.63
204	0.30	-1.7	0.65
261	0.36	-2.6	0.65
235	0.69	-0.9	0.65
266	0.76	-0.6	0.69
275	0.13	-4.3	0.75
118	0.42	-2.1	0.81
224	0.34	-2.6	0.83

APPENDIX D (continued)

TABLE XXIV

STUDENTS WHO TOOK THE HCL SURVEY WHO DISPLAYED *INFORMATIVE-NOISY* RESPONSE PATTERNS

Student #	MNSQ Outfit	MNSQ Infit	Point-Measure Correlation
33	1.43	1.36	-0.10
63	2.00	2.00	-0.21
71	1.34	1.35	-0.16
73	1.54	1.43	-0.14
107	1.47	1.46	-0.04
127	1.89	1.83	-0.06
132	3.02	2.60	-0.72
146	1.75	1.67	-0.02
210	1.32	1.29	-0.08
227	2.42	2.30	-0.13

TABLE XXV

STUDENTS WHO TOOK THE HCL SURVEY WHO DISPLAYED *NON-INFORMATIVE* OR *CONTRADICTORY* RESPONSE PATTERNS

Student #	MNSQ Outfit	MNSQ Infit	Point-Measure Correlation
63	2.00	2.00	-0.21
73	1.54	1.43	-0.14
127	1.89	1.83	-0.06
132	3.02	2.60	-0.72
146	1.75	1.67	-0.02
227	2.42	2.30	-0.13

VITA

Jennifer Sweet

EDUCATION

Doctor of Philosophy in Educational Psychology **2016**
Focus Area: Measurement, Evaluation, Statistics, and Assessment
University of Illinois at Chicago
Dissertation title: "A Model for Improving Survey Outcomes by Reducing Cognitive Load"
Advisor: Carol Myford, PhD

Master of Science in Counseling and Student Personnel **2005**
Oklahoma State University, Stillwater, Oklahoma

Bachelor of Science in Biology **2002**
University of Wisconsin – River Falls, River Falls, Wisconsin

RESEARCH EXPERIENCE

Undergraduate Research Assistant
Departments of Educational Psychology and Biology, **2000-2003**
University of Wisconsin-River Falls

- Contributed to NSF-funded project to explore use of a case-based multimedia tool, Case It!, designed to help students explore biotechnology.
- Conducted and transcribed interviews with students to evaluate effectiveness of the program.

PROFESSIONAL EXPERIENCE

Associate Director, Office for Teaching, Learning & Assessment **2012 – Present**
DePaul University

- Coordinate annual assessment project and reporting process for all academic programs at DePaul University

Assessment Specialist **2007 - 2012**
University of Illinois at Chicago

- Coordinated general education and degree program assessment implementation.
- Provided consultation to faculty and staff regarding development, implementation, and improvement of assessment processes.
- Assisted in the implementation of an online assessment management system, Tk20, to document assessment activities.
- Provided training, consultation, and support to faculty and staff in planning, implementing, and documenting assessment activities in Tk20.

- Analyzed assessment results and generated reports for multiple audiences.

Residence Director for Administration and Assessment **2005-2007**
University of Illinois at Chicago

- Produced and maintained an assessment plan for the department based on mission, vision and goals.
- Coordinated administrative housing processes, including billing, housing assignments and contract cancellations.
- Supervised two graduate assistants in the central housing office.

TEACHING EXPERIENCE

Science in the City **Fall 2014**
DePaul University

- Staff Professional Instructor for Liberal Studies first year course.
- Worked in a teaching team, including a faculty member, a staff member, and a student to teach and assess first year students on content related to the role of science in the city of Chicago and near suburbs.

Summer Smarts Instructor **Summer 2004-2005**
Rondo Community Education
Saint Paul Public Schools, Saint Paul, Minnesota

- Taught science courses for a summer enrichment program for inner city youth
- Planned lessons and activities for students, ranging in age from seven to ten
- Coordinated a week-long summer science camp for children, ages four to eleven

PUBLICATIONS

Bergland, M., Lundeberg, M. A., Klyczek, K., Sweet, J., Emmons, J., Martin, C., Marsh, K., Werner, J., & Jarvis-Uetz, M. (2006). Exploring biotechnology using case-based multimedia. *American Biology Teacher*, *68*, 81-86. doi: 10.1662/0002-7685(2006)068[0081:EBUCM]2.0.CO;2.

PROFESSIONAL PRESENTATIONS

Nelson, J., Jagman, H., & Sweet, J. (2016, August). *Mission possible: Teaching to institutional goals.* International Federation of Library Associations and Institutions satellite preconference workshop, DePaul University, Chicago, IL.

Milligan, S., & Sweet, J. (2016, January). *Faculty development week assessment workshop*. [Invited workshop facilitators]. Kennedy King College, Chicago, IL.

Sweet, J. (2016, June). *Moving your assessment program beyond compliance*. Association for the Assessment of Learning in Higher Education Conference, Milwaukee, WI.

Sweet, J., & Milligan, S. (2016, June). *Responding constructively to criticism in assessment*. Association for the Assessment of Learning in Higher Education Conference, Milwaukee, WI.

Sweet, J., & Milligan, S. (2015, June). *Building agency in assessment through an inter-institutional program.* Association for the Assessment of Learning in Higher Education Conference, Lexington, KY.

Walters, S., Yancey, N., Liou, J., Bartholomew, T., & Sweet, J. (2015, June). *Overcoming challenges in implementing graduate program assessment*. Association for the Assessment of Learning in Higher Education Conference, Lexington, KY.

Chaden, C., & Sweet, J. (2015, June). *Using academic and co-curricular program assessment to assess institutional learning outcomes.* Association for the Assessment of Learning in Higher Education Conference, Lexington, KY.

Milligan, S., & Sweet, J. (2014, September). *How do we know what students are learning?* [Invited workshop facilitators]. Chicago State University, Chicago, IL.

Sweet, J., & Davidson, L. (2013, May). *Understanding the role of academic advising in student learning.* DePaul Academic Advising Network Drive-In Conference, DePaul University, Chicago, IL.

Sweet, J., & Milligan, S. (2013, June). *Integrating assessment across the curricular and co-curricular domains.* Association for the Assessment of Learning in Higher Education Conference, Lexington, KY.

Li, R., & Sweet, J. (2012). *Designing and coordinating a general education program assessment process.* [Invited Presenter]. Association of American Colleges and Universities General Education Conference, Boston, MA.

Li, R., & Sweet, J. (2012). *Designing and coordinating a general education program assessment process: A case study.* [Invited Presenter]. Tk20 Webinar.

Arkalgund, R., Sweet, J., & Rizvi, S. (2012). *An ontological framework for designing the assessment of learning outcomes.* Association for the Assessment of Learning in Higher Education, Albuquerque, NM.

Li, R., & Sweet, J. (2012). *Utilizing institutional strengths to address barriers to general education assessment.* Higher Learning Commission annual meeting, Chicago, IL.

Milligan, S., & Sweet, J. (2011, April). *Improving survey design*. [Invited Lecturer, ELPS 431 Evaluation in Higher Education]. Loyola University Chicago, Chicago, IL.

Sweet, J., & Milligan, S. (2010, October*). Improving assessment through communities of practice*. IUPUI Assessment Institute, Indianapolis, IN.

Huss, R., & Sweet, J. (2005, July). *Freshmen discover, but do not navigate, the waters of suites and apartments.* Association of College and University Housing Officers – International, Milwaukee, WI.

Lundeberg, M., Bergland, M., & Sweet, J. (2002, April). *Using pre-service science teachers as research assistants for Case It!, An international web-based project involving collaborative case-based learning in biology*. Pathways to Change: An International Conference on Transforming Math and Science Education in the K16 Curriculum, Arlington, VA.

AWARDS AND HONORS

Best Paper for 2012 Collection of Papers **April 2012**
Higher Learning Commission

Regional Best Program **February 2005**
Southwest Association of College and University Housing Officers

Chancellor's Award **May 2002**
University of Wisconsin – River Falls