ABSTRACT

ADAMS, MEREDITH JANE DEAN. No Evaluation Left Behind: Nonresponse in Online Course Evaluations. (Under the direction of Paul Umbach).

Technological advances have enabled colleges and universities to administer course evaluations online, forgoing the traditional paper-and-pencil methods. Consequently, many of these institutions have encountered low response rates, and little research is available on this topic. To increase understanding about course evaluation participation in the online environment, this study examined over 22,000 undergraduates to whom the university administered almost 135,000 evaluations. Multilevel models were constructed to analyze the data, nesting course evaluations within students. Several variables emerged as significant predictors of participation, some of which were consistent with previous research and aligned with theories of survey nonresponse. In this study, respondents to SETs were typically female, with a GPA at or above 3.0, over 25 years of age, were first-year or senior students, not African American or Asian, and in a realistic major. However, characteristics of the course also tended to influence participation. Submissions of SETs were more likely if the course and student's major were in the same department; an A, B, or C grade also increased the likelihood of response. Implications for research and practical applications for institutions are also addressed, including ways to combat survey fatigue, increase the salience of the survey, and increase the number of opportunities that a student has to complete course evaluations.

No Evaluation Left Behind: Nonresponse in Online Course Evaluations

by
Meredith Jane Dean Adams

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Educational Research and Policy Analysis

Raleigh, NC

2010

APPROVED BY:

Paul Umbach Chair of Advisory Committee	Gerald Ponder
Betsy Brown	James Bartlett

DEDICATION

To Luna, Reesey, Howl, and Chopsticks

BIOGRAPHY

I was born in Richmond, Virginia and have attended too many schools to count. Prior to attending North Carolina State University, I earned Bachelor's and Master's degrees at Virginia Commonwealth University (VCU '05).

Throughout my postsecondary and graduate education, my work experiences have remained in education but have varied across educational contexts. In primary, secondary, at-risk, and postsecondary institutions, I have taught Spanish, English to speakers of other languages, sheltered-instruction language arts and social studies, and linguistic and cross-cultural communication. While at NC State, I was awarded teaching and research assistantships working with assessment, statistics, and postsecondary teaching; as a research assistant I also worked and conducted research in the Dean's office and in Knowledge Management and Assessment.

While "in school", I've somehow been able to travel a great deal. However, I've been to more countries south of the equator than north of it, so much traveling is still left. When I'm not working or traveling or completing degree requirements, I enjoy playing the guitar, gardening, eating vegan food, participating in almost any outdoor activity, and living simply. But I will never stop learning, researching, and understanding.

ACKNOWLEDGEMENTS

This process has been a team effort, and without these team members, this path to a PhD would not have been as smooth. In fact, I'd probably still be narrowing down my topic. Consequently, there are a lot of people to thank, including friends, family, and colleagues.

I know I have consulted at least half of the College of Education at NC State; this includes faculty, staff, and students. So many of you have helped me throughout this process, including Bonnie Fusarelli, Tamara Young, Malina Monaco, and several others. Moreover, my dissertation group has been so encouraging; I'm still not sure how our group worked with people from disciplines spanning almost every NC State college – but it did. For listening, advising, helping, and providing, I thank everyone who has contributed to this team effort.

I chose my committee for a reason, so I'm glad they agreed to join. Betsy Brown and Gerald Ponder taught the very first class I took – "Teaching in College". The next fall, I was back with them as a teaching assistant for the same course, and they *still* decided to join my dissertation committee. Further along in my coursework, I took two courses each from Paul Umbach and James Bartlett, and they decided to stick around longer, too. (I didn't fail. They were all different courses.) The expertise on this committee is immeasurable, and I am more than proud to have their names on the title page. Paul Umbach, the committee chair, provided great direction and feedback. He's brilliant. I truly appreciate my committee's presence and contributions to this entire process.

Several more individuals have been integral to this process. Valerie Bryant probably will never read this, but she was the first person to really believe in me, accept me for who I am, and encourage me to get this Ph.D. Likewise, it is probable that Susan Adams and Sue Bullard also have no idea how much their support meant. Finally, Brendan deserves to share in the joy of finishing this degree just as much as I. He believes in me more than I believe in myself, and that encouragement went a long way. I have heard that many marriages end ~~because of~~ during Ph.D. programs; on the contrary I could not have done this without Brendan. Even through this no-so-painless process, we still had adventures to the Galápagos Islands last year, to Perú this year, and various other road trips. Most of all I look forward to celebrating our five-year anniversary in December.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER TWO: REVIEW OF THE LITERATURE

CHAPTER THREE: METHODS

CHAPTER FOUR: RESULTS

CHAPTER ONE: INTRODUCTION

It is the last week of classes, and Susan is preparing for her final exams and projects that are due. She takes a break and checks her email. There she reads two emails from friends (what are we doing tonight?), one from her mom (when will you be home for break?), one from her boss (did she finish that order?), another from a classmate (about their final group project), and then a few emails about winter sales at the mall. Then she sees one from her university and opens it. Susan is not sure who exactly sent this email, but they want her to click a link and complete an evaluation for each of her six classes she is taking this semester. Every semester around the same time, Susan receives these emails from her university wanting her to take these surveys, and every semester she makes a decision to ignore or respond to this request.

This end-of-course tradition is familiar to modern-day college students. Nearly every post-secondary institution employs these evaluative surveys to collect information from students' perspectives about their experiences in their courses and about their instructors. These evaluations, increasingly delivered in a web-based format, are completed at the end of each semester or term when the course finishes. Course evaluations measure students' perceptions of effectiveness and quality of the course taught and its instructor. Student evaluations of teaching, also called course evaluations or SETs, have formally existed since the 1920s (Guthrie, 1954; Remmers & Brandenburg, 1927). Instructors receive aggregate results intended to improve their teaching. In addition, administrators may use results from SETs to help determine

promotion and tenure of faculty, as well as to improve courses within the programs. Since their inception, SETs have been controversial (Detchen, 1940; Guthrie, 1953). As an overwhelming majority of universities utilize low-cost web-based SETs, this topic has become even more controversial (Costin, Greenough, & Menges, 1971; Crumbley & Fliedner, 2002; Crumbley & Reichelt, 2009; Haskell, 1997; Langbein, 2005; Stake, 1997). Technology now offers inexpensive methods of notification, delivery, completion, and analysis (Dommeyer, Baum, Hanna, & Chapman, 2004). With these advances throughout the past decade, colleges and universities that have introduced web-based SETs also have seen reduced response rates (Avery, Bryant, Mathios, Kang, & Bell, 2006; Johnson, 2003; Kucsera & Zimmaro, 2008; Layne, DeCristoforo, & McGinty, 1999). A major concern that most universities have about online SETs revolves around the lack of participation. This study seeks to examine participation in SETs, and chapter one introduces the background, purpose, and overview of the dissertation study, demonstrating the importance of this study to university policy and the current research on nonresponse to SETs.

Importance of the Study

What factors contribute to the submission of an SET? What about these students increases their likelihood to ignore or respond to the call to complete the course survey? The purpose of this study was to predict the likelihood of undergraduate student response to online SETs, called ClassEval, at North Carolina State University (NC State), the largest university in the state. This study focused on unit nonresponse, indicating submission of the SET.

It is important to understand unit nonresponse for several reasons. First, nonresponse increases the potential for error and is a threat to external validity. In most cases, survey nonresponse is not random (Dillman, Smyth, & Christian, 2009; Groves, 1989; Porter, 2004a); that is, there is a reason for its occurrence. When respondents and nonrespondents characteristically or descriptively differ, nonresponse error in the form of bias can result. As nonresponse increases, the likelihood that nonrespondents' opinions differ from respondents' also increases (Groves et al., 2004; Groves & Peytcheva, 2008; Porter & Umbach, 2006a; Lindner, Murphy, & Briers, 2001). Therefore, low response rates are likely to increase bias in results, thus affecting the external validity of a study's implications (Groves, 1989; Groves, Dillman, Eltinge, & Little, 2002; Groves et al., 2004; Groves & Peytcheva, 2008; Porter, 2004b; Rubin & Zanutto, 2002; Umbach, 2005). Nonresponse bias is important to acknowledge, since survey results seek to be both generalizeable to the population and representative of what everyone invited to complete the survey believes (Groves et al., 2004). With the survey now in an online format, many campuses suffer from low response rates to their SETs, and quality data for decision-making processes may be unavailable (Avery et al., 2006; Dillman et al., 2009; Groves et al., 2004).

The second reason it is important to study SET unit nonresponse is due to the steady decline in survey participation in recent times (Dillman, Eltinge, Groves, & Little, 2002; Groves & Couper, 1998). At the same time, to decrease costs and facilitate the mode of data collection and analysis, survey administrators are switching from paper to web-based modes of delivery, and response rates have declined even more. This

phenomenon is not only specific to general survey research (Dillman et al., 2009; Groves et al., 2009), but it is also prevalent according to research on survey administration at postsecondary institutions, including SETs (Avery et al., 2006; Dey, 1997; McGourty, Scoles, & Thorpe, 2002a, 2002b; Porter, 2004b; Sax, Gilmartin, & Bryant, 2003; Thorpe, 2002). Due to unstandardized SET instruments and administration modes, paper response rates at NC State are not comparable with their online counterpart. However, the office of institutional research has noted an overall decrease in ClassEval response rates since the implementation of the web-based course evaluation system in 2007 (UPA, 2008b). Nearly every university implements end-of-course SETs, and the savings associated with time and costs have influenced most institutions to place these evaluations in web-based forms. However, web-based surveys have seen an increase in nonresponse.

The third and last major reason to study nonresponse in university SETs comes from their utility in campus administration and faculty work. As nonresponse increases, the potential for error in any survey, including SETs, also increases (Groves et al., 2004). The quality of data and their results tends to weaken when response rates decline (Groves et al., 2004; Groves & Couper, 1998). These evaluations may have multiple uses at various institutions, and there are some important decisions that typically consider the results of SETs. For example, administrators use teaching evaluations for annual review, promotion, tenure, and reappointment decisions. Faculty can also use the results to improve their instruction and apply for grants or awards (Marsh & Roche, 1997; Porter & Whitcomb, 2005). Department heads may consider results from SETs to decide whether

to keep the course in the curriculum or to change it. In the age of data-driven decision-making, it is imperative to collect responses representative of the whole population, but many universities fail in obtaining high response rates from online SETs. Because the results from end-of-the semester SETs can have such high stakes, it is important that we understand the potential for nonresponse error.

This study explored what influences the likelihood of responding to SETs, using data provided by the University Planning and Analysis (UPA) office from the fall semester of 2009. Over 20 characteristics of students and courses analyzed the dependent variable – submission or no submission of the SET. A series of multilevel models examined factors related to the likelihood of response and differences between respondents and nonrespondents.

Background and Theoretical Framework

Since the data collection mode of semester-end teaching evaluations changed from using paper surveys to employing web-based applications, little research has been conducted on nonresponse at various institutions across the U.S. Conflicting evidence debates which mode of collection receives better response, is more efficient, and provides better results. While some researchers suggest that paper surveys have a higher response rate than their online counterparts, others claim that online surveys achieve a higher response rate (Porter, 2004a). Despite the mode of course evaluation data collection, most institutions report lowered response rates with the implementation of web-based SETs (Avery et al., 2006; McGourty et al., 2002; Sax et al., 2003; Thorpe, 2002).

Before the technological advances in the method of SET administration, nonresponse error was not such a large concern for universities; instead, discussions of coverage error were more relevant. It was possible that not all students had the opportunity to respond, or ineligible responses were collected. At NC State, extra survey forms included in the course evaluation packet may be responsible for yielding 0-200% response rates (UPA, 2008c). In 2007, the mode of delivery changed to an online format, and the university's institutional research office has measured relatively consistent response rates at 54-59%, though these rates fall each semester (UPA, 2008b). Moreover, this study discovered that SETs for Fall 2009 classes yielded a 48.1% response rate – the lowest since the online implementation.

What characteristics of students and their courses distinguish respondents from nonrespondents? When studying how and if participants respond, survey research reveals significant differences among types of people (Goyder, Warriner, & Miller, 2002; Moore & Tarnai, 2002; Porter & Whitcomb, 2005). In studies conducted on SETs, student ratings have revealed bias. For example, high or low ratings have been found to correlate with gender, cultural background (Davis, Hirschberg, Lye, & Johnston, 2007), student performance (Cohen, 1981; Costin, 1978; Crumbley & Reichelt, 2009; Isely & Singh, 2005; Marsh, 2007; McKeachie 1969), and year in school (Davis et al., 2007; McKeachie, 1979). If one were to find a relationship between these characteristics and the likelihood of response, it would seem reasonable to interpret the data with caution.

Relatively recent literature on participation in web-based evaluations and surveys also seems to demonstrate differences in the likelihood of response among students at

universities. For example, females are more likely to respond than males (Avery et al., 2006; Dey, 1997; Fidelman, 2007; McGourty et al., 2002a, 2002b; Porter, 2004b; Porter & Umbach, 2006a; Porter & Whitcomb, 2005; Sax, Gilmartin, Lee, & Hagedorn, 2008). High performance and achievement (as measured by grade, cumulative GPA, and SAT score) are positively related to the likelihood of response (Avery et al., 2006; Fidelman, 2007; McGourty et al., 2002a, 2002b; Porter & Umbach, 2006a; Porter & Whitcomb, 2005). Lastly, white students generally respond at higher rates than students of color (Avery et al., 2006; Clarksberg, Robertson, & Einarson, 2008; Dey, 1997; Fidelman, 2007; Porter & Umbach, 2006a). In a study of the National Survey of Student Engagement (NSSE), a survey of postsecondary undergraduate students, Porter and Umbach (2006a) found that students with lower SAT scores, male students, and students of color were less likely to respond. A review of the literature found in chapter two addresses additional background research on this topic.

Theories of Survey Participation

A review of the literature and theories may help explain nonresponse and serves as the framework further explored in the next chapter. Three main theories aided the developing understanding influences on survey nonresponse in this study. The first is social exchange. This theory originally described all interactions between people as either giving or taking, and that such actions seek to be reciprocated (Blau, 1964). However, Dillman (1978), describing survey participation as a social activity, saw social exchange theory as beneficial in explaining nonresponse in surveys (see also Dillman et al., 2009). Choosing to respond may occur due to feelings of duty or wanting to help someone or the

institution. For example, college teachers may convey the idea that SETs mean a lot to them and that their students' opinions truly matter to them. Students may also respond at higher rates to SETs associated with courses in their academic environment. The sense of duty students felt may have increased response rates, the reciprocal “giving back” for services rendered by the institution (Blau, 1964). Nonresponse rates may also be explained by personal culture. Patterns of social participation such as minority group oppression, social distance perceptions, and helping behaviors have fielded discussion and results in nonresponse survey research (Johnson, O’Rourke, Burris, & Owens, 2002).

Second, a theory of academic disciplines, based on Holland’s theory of careers (Holland, 1966, 1997; Smart, Feldman, & Ethington, 2000), also guided the analysis. Originally, Holland demonstrated that people choose careers that match their personality. He characterized people based on six types: realistic, investigative, artistic, social, enterprising, and conventional. Realistic types avoid interpersonal interactions and prefer activities with predictable outputs. Investigative people seek to understand and control their surroundings and avoid social activities. Artistic people do not prefer order and routine, but instead look for ways to express themselves creatively. The social personality type avoids ordered activities with predictable outputs and a higher capacity for human interaction. Enterprising types of people are extroverted and very politically oriented. They manipulate to attain their goals, and are unscientific. Lastly, conventional people enjoy structure and place significant priority on financial achievements and power.

Smart, Feldman, and Ethington (2000) applied Holland’s theory of career choice to academic fields of study. Whereas Holland theorized that certain personality and

characteristic traits influence vocational choice, these authors theorized that the same traits also influence a postsecondary student's choice of major and academic environment. Smart et al. argued not only that the same six personality classifications align with the academic field in which a student majors, but also that the academic environment encourages certain behaviors and attitudes for those within their community. Using personality as measures of if and how students respond, some studies have demonstrated potential inherent predispositions, typically extroverts and social students (Marcus & Schutz, 2005; Sax et al., 2003; Yu, Jannasch-Pennell, DiGangi, Kim, & Andrews, 2007). It may be possible that personality types measured by major can also predict nonresponse. For example, Porter and Whitcomb (2005) found that the decision to respond to the National Survey of Student Engagement (NSSE) correlated with the major environment type of the student (Holland, 1997; Rosen, Holmberg, & Holland, 1997; Smart et al., 2000). Artistic majors were less likely to submit surveys, and students within investigative majors were more inclined to respond. This study of nonresponse to SETs at NC State also examined the six environments associated with student majors.

Finally, the theory of leverage-salience (Groves, Singer, & Corning, 2000) explains that a person's decision to cooperate consists of multiple aspects unique to each individual. The survey design, a person's situation, and other variables combined can influence an individual's decision to participate in the survey. However, for the same or similar survey design, situation, and variables, another person may decline the survey invitation. Reactions and decisions are never the same, and each individual's situation and rationale affects survey participation differently. This theory represents a holistic,

contextual, and individualized explanation of survey participation. Several variables, characteristics, and attributes may influence different people in different ways, appealing to the uniqueness and individuality of the context. The influences on the decision to participate depend on circumstances related to each individual, attributes of survey design, the survey topic, and personal preferences. In studies of SET nonresponse, it is possible that attributes of SETs influencing individuals to respond also influence particular populations to respond. The potential for this bias may become clearer in studying the characteristics of respondents and nonrespondents.

These approaches are helpful in studying nonresponse in SETs at universities. It is important to note that variables other than major may also increase the salience, or attraction, of the SETs to each respondent. Other theories such as survey fatigue (Porter, Whitcomb, & Weitzer, 2004), survey saliency (Groves et al., 2004), and opportunity costs (Dillman et al., 2002; Groves et al., 2004) also informed the study and explained the decision to participate. The review of the literature found in the next chapter explores additional theories and the potential for nonresponse in greater detail.

Purpose and Analysis

The purpose of this study is to investigate student nonresponse in end-of-semester SETs. More specifically, this study explored the factors and influences of unit nonresponse for all undergraduates across NC State. This is the first comprehensive study of nonresponse to web-based SETs, because most studies are either obsolete or have examined a handful of classes (Ho & Shapiro, 2008; McGourty et al., 2002a; McGourty et al., 2002b; Thorpe, 2002; Vehovar, Batagelj, Manfreda, & Zaletel, 2002). Other studies

have taken place at institutions unlike the current university, have not used as many variables as the dataset for this study, and have excluded potentially important variables (Ho & Shapiro, 2008; Fidelman, 2007; Jones, 2009). Moreover, much is at stake when considering nonrespondents in this context; administrators consider SET results in the process of making personnel and academic decisions. To reduce the risk for nonresponse error and improve data quality, uncovering factors related to the probability of nonresponse is a primary concern for institutions implementing online SETs. The following research questions guided the current study of unit nonresponse:

1. To what extent was the likelihood of response to SETs related to student-level measures, such as demographics (e.g. age, gender, race) or academic performance (e.g. GPA, SAT score, grade in course)?
2. To what extent was the likelihood of response related to personality and academic environment as described by Holland (1973; 1997)?
3. To what extent did salience (e.g. if course is in the same department as the student's major, student grade in course) increase the likelihood of response? Did the relationship vary significantly between students? Could student-level measures, such as major and gender, explain this variation?
4. To what extent did survey fatigue (e.g. number of SETs to complete, class rank) predict response rates? Did response rates correlate to opportunity costs (e.g. number of credit hours, number of courses taken that semester)? To what extent

did culture (e.g. environment, institutional engagement, personal background) vary response rates?

Over 20 independent variables analyzed the dependent variable of unit submission: did the student submit the SET? These variables include the number of credit hours the student took, the department that hosted or sponsored the course, student's major, GPA, gender, nationality, number of credits, athletic participation, and transfer status. Data analysis focused on a linear mixed model construction called hierarchical linear modeling (HLM) or multilevel model (MLM). A multilevel model is useful in this study because the observations are not independent (Garson, 2009; Heck & Thomas, 2000; West, Welch, & Galecki, 2007). In MLM, data are naturally nested or clustered within groups (Allison, 1999; Raudenbush & Bryk, 2002). In this study, clustering exists at two main levels – the survey level (Level 1) and the student level (Level 2). The survey level contains variables related to characteristics of the evaluation requested for completion. This dataset contained 134,929 cases; each evaluation was one case. Therefore, if a student took five classes, there are five cases attributed to that student. The second level includes variables related to students and their characteristics, some of which are GPA, number of credits taken that semester, gender, and nationality.

Educational research commonly uses HLM models because of this clustering and lack of independent observations. These models are also preferable over typical regression models because HLM provides more precise point estimates and standard

errors than regression models when analyzing nested data structures (Allison, 1999; Uekawa, 2005).

The results from HLM modeling analysis relied on an encompassing framework that includes theories as well as an in-depth review of the literature. Concepts and theories such as social exchange (Blau, 1964; Dillman, 1978), topic salience (Groves et al., 2004), and academic disciplines (Holland, 1973, 1997; Smart et al., 1999, 2000) guided the model specification to fulfill the purpose of this study.

Overview of the Study

This study of unit nonresponse in SETs focused on undergraduates at NC State. Because it is unlike any study on SETs, its comprehensive nature illuminates a gap in the research on nonresponse to SETs and begins to address it. There are several differences between this and previous studies on SETs. First, until this time most research into student evaluations of teaching has debated their accuracy, effectiveness, usefulness, and modes of administration, (Avery et al., 2006; Feldman, 2007; Marsh, 2007; Marsh & Roche, 1997). Second, other studies have considered nonresponse in only a handful of classes (Fidelman, 2007; Ho & Shapiro, 2008; Jones, 2009; McGourty et al., 2002a, 2002b; Thorpe, 2002). Additionally, these and other studies have taken place at smaller or private institutions, did not use as many potential factors of nonresponse, and excluded potentially important variables (Fidelman, 2007; Ho & Shapiro, 2008; Jones, 2009). By including all students and having a larger amount of variables available, this study is preeminent and more useful in the application of results.

The next chapter examines the literature surrounding the topic of survey participation, where the need for this study becomes more evident. The third chapter of the dissertation proposal describes the data and how they were collected, measured, analyzed, and reported. The methodology chapter will also further describe variables and the process of analyzing the data. The fourth chapter presents results from the analyses, and the final chapter addresses the results and how they answered the research questions and contributed to the theories presented on survey participation. This fifth chapter also contains implications for campuses and methods that will reinforce a campus climate of survey participation.

CHAPTER TWO: REVIEW OF THE LITERATURE

This study sought to predict the likelihood of response to university student evaluations of teaching (SETs) based on student and course characteristics. Survey nonresponse is on the rise, a phenomenon that increases the potential for nonresponse error (Dillman et al., 2002; Groves & Couper, 1998). Error tends to negatively affect data quality and implications of findings, and online surveys seem to exacerbate nonresponse (Berinsky, 2008; Dillman et al., 2002; Groves, 1989; Groves et al., 2002; Groves & Peytcheva, 2008; Porter, 2004b; Rubin & Zanutto, 2002; Umbach, 2005). Because online surveys have time-efficient and cost-effective benefits, postsecondary institutions are increasingly administering web-based instruments. However, many institutions with online SETs have become victims of decreasing response rates (Avery et al., 2006; Dey, 1997; McGourty et al., 2002a, 2002b; Porter, 2004b; Sax et al., 2003; Thorpe, 2002; Umbach, 2004).

This study took place at North Carolina State University (NC State). After the Spring 2007 implementation of ClassEval, NC State's system of web-based class evaluations, the campus institutional research office has seen response rates decline every semester. Error associated with nonresponse is of concern in the postsecondary context because of the utilization of course evaluation results. With these results, instructors have the opportunity to revise and improve their teaching, and administrators conduct annual reviews and make reappointment, promotion, and tenure decisions after reviewing the results. In turn, results also influence salary and other awards or recognition. To address

appropriately nonresponse to online SETs, campuses must understand influences of nonresponse. To discover what factors influenced the likelihood of response to SETs at NC State, the following research questions guided the current study:

1. To what extent was the likelihood of response to SETs related to student-level measures, such as demographics (e.g. age, gender, race) or academic performance (e.g. GPA, SAT score, grade in course)?
2. To what extent was the likelihood of response related to personality and academic environment (Rosen et al., 1997; Holland, 1973; Smart et al., 2000)?
3. To what extent did salience (e.g. if course is in the same department as the student's major, student grade in course) increase the likelihood of response? Did the relationship vary significantly between students? Could student-level measures, such as major and gender, explain this variation?
4. To what extent did survey fatigue (e.g. number of SETs to complete, class rank) predict response rates? Did response rates correlate to opportunity costs (e.g. number of credit hours, number of courses taken that semester)? To what extent did culture (e.g. environment, institutional engagement, personal background) vary response rates?

This literature review examines several topics surrounding these issues. First, the history, purpose, and controversies regarding end-of-course SETs provide a context for the topic. Second, the section on survey participation addresses causes and issues of

nonresponse, and ways to combat it. Theories of survey participation included in the theoretical framework for this study and nonresponse studies specific to university student surveys are also in this section. Finally, an explanation of the ClassEval system and results from analyses conducted by the institutional research office concludes this chapter.

A History of University Course Evaluations

In the 1920s, researchers began to see the need for colleges to know their own students' opinions (Brandenburg & Remmers, 1927; Guthrie, 1954; Remmers & Brandenburg, 1927). Hermann Remmers began to publish multiple studies and their results in the 1920s, becoming perhaps the first consistent and active contributor to the field. The Purdue Rating Scale for Instruction (PRSI), implemented at Purdue University, measured teaching effectiveness for several decades. Remmers and his colleagues demonstrated that this process of evaluating teaching demonstrated the reliability and validity of teaching evaluations (Brandenburg & Remmers, 1927; Heilman & Armentrout, 1936; Remmers, 1928, 1930, 1934; Remmers & Brandenburg, 1927). Even in the formative years, such studies continued to support validity and reliability of SETs (Guthrie, 1953; Root, 1931). Even as early as 1925, controversy over the rating and judging of college instructors existed (Detchen, 1940; Guthrie, 1953; Root, 1931). However, researchers continued to provide evidence that students are able to give a valid opinion and that instruction can improve from teaching evaluation results (Detchen, 1940; Drucker & Remmers, 1951; Guthrie, 1954; Remmers, 1934; Stalnaker & Remmers, 1928).

Methods of evaluating instructors became prevalent themes in research through the 1950s and 1960s, and administrators increased their use of end-of-course SETs to assess faculty performance (Astin & Lee, 1967; Centra, 1977; Costin et al., 1971; Gustad, 1961, 1967). As baby boomers began their postsecondary studies, the demand for faculty increased (Seldin, 1975). Consequently, the need for instructors was difficult to meet and teaching quality decreased. Researchers began to recognize that teaching is not one-dimensional (Barr et al., 1953). Reliability and validity studies continued to demonstrate that students' evaluations were stable, consistent, and useful (Guthrie, 1954; Miklich, 1969; Scates, Baker, & Remmers, 1951). The literature during this time described and evaluated the multiple methods used for evaluating instructors other than SETs (Gustad, 1961; Astin & Lee, 1967; Seldin, 1975). Most research further validated evaluations of teaching, as alumni opinions, peer evaluations, and other sources revealed strong correlations (Cashin, 1994; Seldin, 1984).

Consistent evaluation of instruction became more widespread in the 1970s and 1980s as quality of education became the focus. In a period Seldin called the "anxious years for institutions of higher education" (1984, p.6), enrollment began to sharply decline as baby-boomers left colleges and universities. In "the hunt for more money" (p.11), private companies operating their enterprises under the consumer and customer models increased their ties and investments in particular degree programs.

In light of the enrollment decrease, institutions were able to focus more on the quality of teaching, which would also attract more students, increase retention rates, and make the institution more competitive. SETs were able to facilitate this process,

especially because these opinions came directly from the consumers of the course. Researchers conducted studies on institution-wide teaching evaluation instruments, focusing on understanding the best way to measure teaching effectiveness. Several major players emerged. Cashin and Perrin (1978) studied the IDEA instrument administered to students at Kansas State University between 1975 and 1978. Cohen's (1981) evaluation instrument measured eight dimensions of teaching effectiveness. The Students' Evaluations of Educational Quality (SEEQ) (Marsh, 1984) measured several dimensions of teaching aspects. Feldman's research pointed to 28 different measures of teaching effectiveness (1976a, 1976b, 1989).

Before the technological advances in data collection of SETs, nonresponse error was not a large issue for universities; discussions of coverage error were more relevant. For example, in paper format, typically all class members that were present on the day the instrument was administered would complete their SETs. In this case the sampling frame failed to include some of the target population because it did not count the opinions of those absent, a problem of undercoverage (Groves et al., 2004). Overcoverage may also have occurred when students completed more than one evaluation per class or when ineligible persons participated. The current trends in research, administration, and evaluation of student evaluations of teaching revolve around the mode of data collection. In the last decade, most postsecondary institutions have decided to cease using paper surveys and turn to a web-based system. With the implementation of this technology, most campuses have seen a decrease in response rates (Avery et al., 2006; McGourty et al., 2002a, 2002b; Sax et al., 2003; Thorpe, 2002). Additionally, while response rates to

surveys in general continue to decline (Dillman et al., 2002; Groves & Couper, 1998), the potential increases for error and low data quality (Berinsky, 2008; Dillman et al., 2002; Groves, 1989; Groves et al., 2002; Groves & Peytcheva, 2008; Porter, 2004b; Rubin & Zanutto, 2002; Umbach, 2005).

The Importance of SETs and the Need for Unbiased Data

Since the 1920s, the major purpose of assessing student perspectives on courses has been to improve instructors' teaching quality (Detchen, 1940; Guthrie, 1953; Wilson, 1932). Not too much has changed in the present-day institution. While improving instruction is still one of the major reasons evaluations of teaching exist, modern-day SETs have incorporated additional purposes. After realizing these purposes and the utility of the results in higher education in this section, the need becomes clearer to increase response rates.

Hebert Marsh (2007) and Kenneth Feldman (2007) give several major reasons why educational institutions collect student evaluation data. Not all are applicable to every institution, but they are, inclusively, a set of potential purposes for collecting data. First, student evaluations of courses and instructors give feedback to instructors to improve their teaching. In the current process of student evaluation data collection, the person or office in charge of the data typically gives the feedback to each instructor after the semester or quarter ends. Marsh states that this first reason for SETs is almost universal to all institutions, but that all may not employ the remaining purposes. Second, personnel decisions also use results as measures of teaching effectiveness. That is, annual reviews, promotion, retention, and tenure decisions may take into account the results of

these evaluations. Many institutions also require that applicants for faculty positions submit their student evaluations of teaching (Walker, Golde, Jones, Bueschel, & Hutchings, 2008). Third, if published, students can view the results to determine with whom or if they should take the course (Feldman, 2007). Fourth, SETs can be a component in quality assurance. Considering the results, department heads and other administrators make curricular and program decisions. Lastly, research on teaching in higher education utilizes results of SETs (Feldman, 2007). Similarly, in this research study I utilized SETs, but only by submission. Participation, not actual student responses, was the focus.

Specific to this study, the UPA office at NC State provides four major purposes for the results of SETs, or ClassEval (2008d). First, department heads use the information for annual review, to increase salary, and assign courses to instructors in their departments. Second, committees and other administrators consider the results of student evaluations when making personnel decisions in tenure and promotion. Faculty members receive the results of the SETs from the courses they taught that semester, from which they can learn how to better their instruction and make adjustments accordingly for the next semester. Lastly, instructors can use SET results to apply for certain grants or awards.

Understanding the high stakes associated with using course evaluation results, it is no wonder that concern over nonresponse exists (Miklich, 1969; Thorpe, 2002). When thousands of students do not complete their end-of-semester course surveys, instructors, department heads, and campus administrators should recognize the need to increase

response rates. In turn, the risk for nonresponse error decreases, and the potential for high-quality results increases.

Establishing the Place of SETs at the Institution

Several issues associated with student evaluations of teaching typically arise when addressing the topic of SETs. While the overarching purpose of students' evaluating courses is to improve the quality of education, many people take issue with three major areas: 1) the students' expertise, attitudes, and manner in which they complete their evaluations, 2) the effectiveness of the instrument including its delivery and results, and 3) the usefulness of the process and its results, including validity and reliability. Moreover, when careers are potentially at stake, defensive attitudes and behaviors are inevitable. A great deal of controversy about the evaluations of college courses from the perspective of the student has existed since their inception. The following sections explain that SETs are helpful and necessary tools of the university administration and faculty members. Student opinions are not optional; the results are effective, and the process is valid, reliable, and useful.

Students

The first of the three major concerns when addressing SETs are the students. Good teachers are aware of their audience and adjust instruction accordingly (Feldman, 1989, 2000; Nilson, 2003; Seldin, 1999). It is out of concern for learning that teachers adapt to their students by seeking to understand what their students need to learn and responding to those needs. However, authors have expressed their doubts about students' qualifications to judge instruction, their morality, and their roles as consumers in the

institutions (Crumbley & Fliedner, 2002; Haskell, 1997; Crumbley, & Reichelt, 2009; Langbein, 2005; Stake, 1997). Stake wrote that the SETs are a symbol that the university has no faith in the person they hire. However, both the age of accountability and modern institutional practices dictate that students are indeed consumers of education and the services provided by colleges and universities (Costin et al., 1971; Lechuga, 2008; Slaughter & Rhoades, 2004). It is therefore generally agreed that evaluations of courses by students will continue to be a presence in higher education (Algozzine et al., 2004; Langbein, 2005; Scott, 1999; Seldin, 2006).

Student attributes can significantly affect how they rate their instructors and courses. Students are increasingly diverse, and many times higher or lower ratings are associated with their different characteristics. These characteristics can include gender, cultural/ethnic background (Davis et al., 2007), student performance and grades (Cohen, 1981; Costin, 1978; Crumbley & Reichelt, 2009; Isely & Singh, 2005; Marsh, 2007; McKeachie 1969), and year in school (Davis et al., 2007; McKeachie, 1979). Additionally, student ratings may be related to student motivation (Cashin, 1994; Marsh, 1984; Marsh, 1987). Two sections found later in this literature review, validity of SETs and nonresponse error, further address potential influences and bias in ratings and response.

Effectiveness of Delivery Method and Results

Utilization.

The second major concern about students' evaluations is their effectiveness. This includes the effectiveness of the delivery method, effectiveness of the results, and how

the process meets its purpose. The two major purposes of SETs are 1) for teachers to improve their instruction, and 2) for personnel decisions (promotion, tenure, annual review, and appointment). The actual usage of these two purposes is difficult to track. Typically, the extent to which faculty and administrators use results remains mostly anecdotal (Cashin, 1994; Cohen, 1990; Marsh 1987; Marsh, 2007; Schmelkin, Spencer, & Gellman, 1997).

There are some possible exceptions to the anecdotal evidence, but there are mixed results. In some longitudinal studies, rates consistently declined throughout the instructor's career (Feldman, 1983; Marsh 2007; Marsh & Dunkin, 1992; Marsh & Hocevar, 1991a, 1991b; Renaud & Murray, 1996; Rotem & Glasman, 1979). Lacking in these studies is the ability to measure if and how instructors actually used the results. Perhaps institutional effects (the expectations of faculty, the manner in which certain institutions view SETs, location) could explain some of the variance discovered by comparing different studies (McGourty et al., 2002a, 2002b; Porter & Umbach, 2006a). Instructors have reported that they find SETs helpful (Schmelkin et al., 1997). Moreover, if they demonstrate an attempt at understanding and change, faculty members are more likely to use their ratings in a meaningful way to improve instruction and ratings (Marsh & Hocevar, 1991b; Pambookian, 1974; Stevens & Aleamoni, 1984).

Data Collection.

Until computers and technology allowed for change in the delivery method of student evaluations of teaching, students completed SETs on paper. Analysis of data was completed by hand, and results were typed and given to instructors (Detchen, 1940). As

technology advanced, optical scanning devices made the analysis of data from these paper evaluations easier (Avery et al., 2006). Computers quickly migrated to work and everyday life, and with the Internet, survey researchers have an increased capacity for contact success. However, different opportunities for survey error arrive with a new mode of collection. With end-of-semester evaluations conducted with paper and pencils, nonresponse error was not a large issue for postsecondary institutions; instead, coverage error was more relevant. For example, in paper format, typically all class members that were present would complete their SETs. In this case the sampling frame failed to include some of the target population because it did not count the opinions of those absent, a problem of undercoverage (Groves et al., 2004). Overcoverage may also have occurred when students completed more than one evaluation per class or when ineligible persons participated.

There are several reasons why the implementation of web-based SETs increased from almost none to almost all over the past decade (Hmieleski & Champagne, 2000). First, paper usage is more expensive. Second, analysis of data and results can be delivered more quickly to the receivers (faculty and administration). Third, electronic surveys provide more security and anonymity for participants. With paper forms, the potential to link handwriting back to students exists. Also, students are typically the carriers and collectors of the paper forms, which may affect anonymity and the security of data.

Much of the institutional and collegial discourse regarding SET response rates varies. Studies have demonstrated that institutional response rates have decreased since

the implementation of web-based evaluations (Avery et al., 2006; Johnson, 2003; Kucsera & Zimmaro, 2008; Layne et al., 1999). However, the mean ratings students assigned to their instructors exhibited no significant difference when comparing paper and web-based instruments (Dommeyer et al., 2004; Gamliel & Davidovitz, 2005; McGhee & Lowell, 2003; Sax et al., 2008; Thorpe, 2002). Despite lower response rates, if the mean ratings for faculty remain consistent, the web-based method is potentially just as effective and reliable as a paper version (Groves, 2006).

Utility of SETs

The final of the three issues most often discussed in the literature is the survey's usefulness. There are three major concerns, including reliability and validity, around which this issue of utility lies: if the process and results are usable, if recipients can utilize the results, and how they use the results. Consideration of the usefulness of the survey can first take place at the survey design level. Every institution (or every department, college or school) typically has self-developed their own instrument for course evaluation use in some way. They may incorporate questions from other sources as well as questions that a committee or institutional research office deemed appropriate. The survey instrument should contain multiple dimensions of teaching to make the survey useful to the instructor and administration (Cashin, 1994; Marsh, 1984; Feldman, 1989) (see next section for more on national SET survey instruments). This builds the foundation for valid and reliable results, based on which instructors can reflect upon and possibly alter their teaching methods.

Reliability.

Marsh and others labeled reliability as a prerequisite for validity (Marsh & Overall, 1979; Marsh, Overall, & Kesler, 1979). Early research on SETs demonstrated high reliability within evaluations of teaching (Guthrie, 1954). Since the 1970s, discussions around SETs have demonstrated that a survey's reliability is much less controversial than its counterpart, validity (Gage, 1974; Rotem & Glasman, 1979). For the purpose of this topic, measures of reliability typically revolve around three major areas: generalizeability, consistency, and stability (Cashin, 1994; Marsh & Roche, 1997).

First, for results to be generalizeable, the results must accurately reflect a) what the students generally think, and b) how the instructor generally teaches. In other words, results must be generalizeable to instructors in all classes they teach. For example, Marsh (1982) found that instructors' ratings did not change significantly when teaching a different course. Gillmore, Kane, and Naccarato (1978) determined that for an increase in the generalizeability of results, as many SETs as possible should be taken into consideration.

Second, the consistency aspect of reliability depends on the raters or participants. Consistency increases when different students in the same course generally agree (Cashin, 1994; Marsh & Roche, 1997). If student ratings are perfectly reliable, then each student would agree to the effectiveness of the instructor, and the results would manifest the level of effectiveness of teaching (Kulik, 2001). However, as Detchen stated 60 years prior (1940), students are different. They have different needs, different characteristics, different temperaments, and different experiences in the classroom. Because students

have different learning styles, they may not perceive teaching effectiveness the same as others in their class. Due to these differences, student responses on the SETs may disagree, potentially decreasing the reliability of the results. Because both generalizeability and consistency tend to increase as the number of raters increases (Cashin, 1994; Gillmore et al., 1978; Marsh, 2007), reviewers of course evaluation results should ensure there is enough information to make an informed decision or conclusion. That is, they should be more aware of response rates and how to combat their decline.

A final aspect of reliability, stability, deals with results over time. Responses over time to the same teacher aid in the measurement of how stable the results are. Unfortunately, several studies demonstrate that teachers' ratings systematically decrease over time (Feldman, 1983; Marsh & Dunkin, 1992; Renaud & Murray, 1996). In a study conducted over a period of 13 years, Marsh and Hocevar (1991b) used ratings from the SEEQ instrument (Students' Evaluations of Educational Quality) (see also Marsh, 1984). For 195 instructors, they found that ratings did not decrease as previous studies discovered. The mean ratings instead showed no systematic changes relating to number of years taught (Marsh, 2007). Credit may be given to the instrument's structure, thereby considering it a reliable and stable instrument.

Validity.

Validity seeks to determine if what is supposed to be measured, in this case courses and instructors, is really being measured. Cashin (1994) described four aspects of validity for SETs. First, he uses student learning as a tool to establish validity. Student learning is "the most widely accepted criterion of effective teaching" (Marsh, 2007,

p.338). Using scores on an external exam (common across all sections developed by an outside party) as a measurement for student learning, Cohen (1981) discovered a high correlation between higher grades on the exam and higher ratings for the instructor. If students learn the most from teachers with higher ratings on their evaluations (Kulik, 2001; Marsh, 2007), then the instrument is more valid.

The second tool used to establish validity is the instructor's self-ratings. Instructor's self-ratings were either significantly or highly correlated with the ratings of their students (Cashin, 1994; Marsh, 1984). Furthermore, moving to the third aspect of validity, student ratings also agree with ratings of others. This includes administrators, alumni, colleagues and other observers, and verbal comments heard outside of the classroom (Blackburn & Clark, 1975; Kulik, 2001; Kulik & McKeachie, 1975; Marsh & Overall, 1980). Such agreement supports the validity of end-of-SETs.

The fourth and final aspect of validity is bias. Bias occurs when responses are different and correlate among certain populations or groups (Groves, 2006; Rogelberg & Luong, 1998). That is, bias occurs when people with a matching characteristic rate similarly, and this rating is different from those outside of that group. Though there is some controversy over what characteristics matter, when they matter, and how to control for bias, research demonstrates that student characteristics, at the very least, minimally affect how students rate their instructors and courses. These characteristics can include gender, cultural/ethnic background, student motivation, faculty rank, faculty expressiveness, course level, academic discipline of the course, difficulty and workload of the course, lack of anonymity, the presence of the instructor in the room, student

performance and grades, the purpose of the results of the ratings, and the student's year in school (Cashin, 1994; Cohen, 1981; Costin, 1978; Crumbley & Reichelt, 2009; Davis et al., 2007; Feldman, 1976a, 1976b; Isely & Singh, 2005; Marsh, 1984, 2007; McKeachie, 1969, 1979). Controlling for some bias and expanding the analysis may decrease these associations and establish an effective and useful process (Centra, 2003; Clayson, 2009; Cohen, 1981; Feldman, 1976a, 1976b, 1989; Marsh & Overall, 1979; Marsh et al., 1979; Marsh & Overall, 1980).

Studies on evaluations of teaching have been controversial since the beginning. Thirty years after the advent of SETs, Remmers and colleagues revealed their distress over the continued arguments (Baker & Remmers, 1952; Remmers & Gage, 1949). They noted that the critics did not take part in continued research to contribute to the field, but instead voiced their objection and dissatisfaction through philosophy. Many studies had already established utility, reliability, and validity (Brandenburg & Remmers, 1927; Heilman & Armentrout, 1936; Remmers, 1928, 1930, 1934; Remmers & Brandenburg, 1927). The literature appears cyclical, and controversies are again finding their way into the literature. However, readers can probably "find individual studies that support almost any conclusion" (Cashin, 1994, p.538). As with any researchable topic, it is important to look at the field of literature as a whole, being able to recognize what does not agree. SETs are generally "reliable, valid, and relatively free from bias, probably more so than any other data used for faculty evaluation… We should not confuse a source of data with the evaluators who use the data to make a judgment" (p.538).

Many authors with serious opposition to the SET process do not regularly contribute to SET research (Baker & Remmers, 1952; Theall, 1997). For example, Larry Crumbley is in accounting at Louisiana State University, Laura Langbein is in public administration at American University, Jeffrey Stake is in law, and Robert Haskell is a professor emeritus in psychology focusing on unconscious cognition and transfer of learning. Additionally, researchers like Feldman, Marsh, Cashin, Kulik, and McKeachie who accept SETs as effective and useful (while acknowledging potential issues) have steadily written on the topics for four or more decades. It has been difficult to find opposing viewpoints as dedicated to the topic as they.

Decades of scholarship have determined that SETs exist for a reason and are valid, reliable, and useful in the evaluation of teaching process. However, if bias has been consistently evident in how students rate their instructors, could biases also exist for students who do not rate their instructors? Are certain populations or characteristics common among nonrespondents and could course characteristics also influence participation? This study focused on these questions. The next section continues the discussion about survey participation and variance in response rates.

Survey Participation: Response or Nonresponse?

Nonresponse occurs when a participant does not complete all or part of the survey. This study focused on unit nonresponse, if submission of the survey was successful, and did not examine partial completion or item analysis. Because the data provided by survey results drive decision-making processes, nonresponse deserves considerable attention (Berinsky, 2008; Groves et al., 2004). Much of the research on

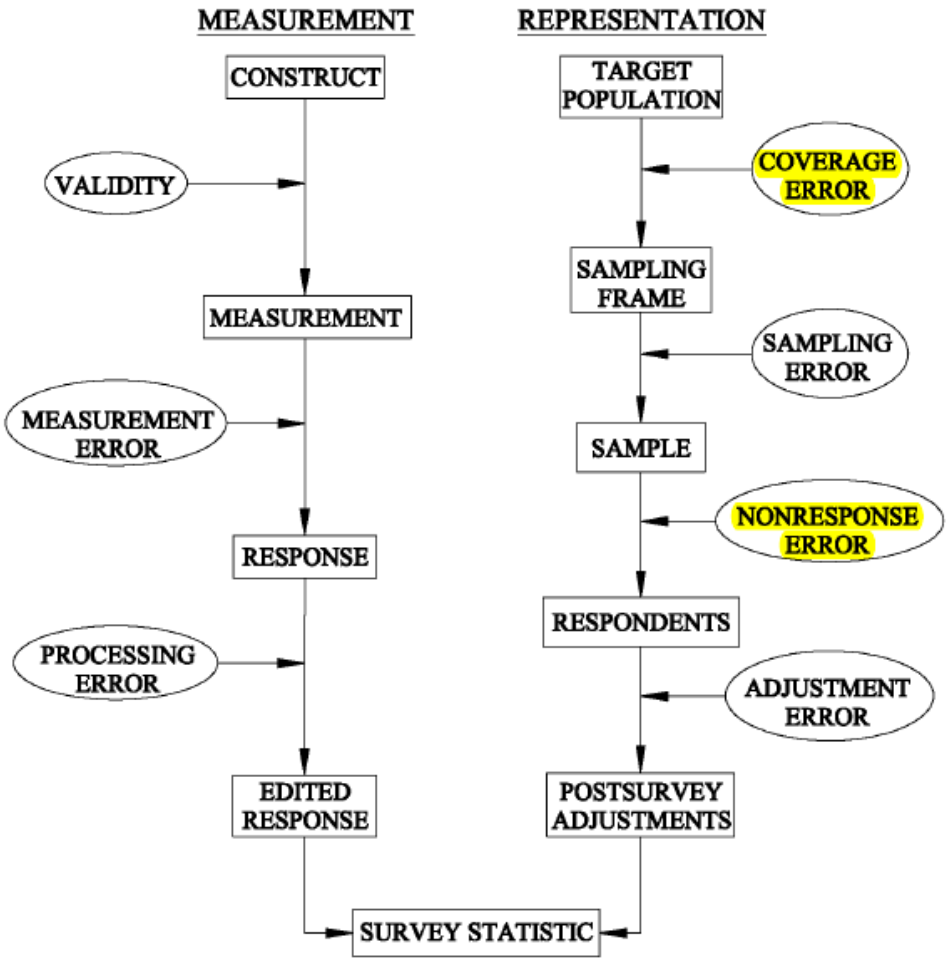
SETs has focused on response rates by method of administration. Perhaps this is due to recent migration of survey mode from paper to web-based instruments (Dillman et al., 2002; Groves et al., 2004). In the few studies particular to SETs, this migration generally reports decreased response rates (Avery et al., 2006; McGourty et al., 2002a, 2002b; Sax et al., 2003, Thorpe, 2002). Because of the high stakes that results of SETs can have, there is a large need to identify factors of and influences on nonresponse.

Until this point, the review of the literature has revolved mostly around responses to evaluations of courses. Eighty-five years of research described how participants view their course and instructor, what students respond, how they respond, if and how the results are utilized, and if the process is effective. However, only briefly mentioned are those who do not complete their SETs at the end of the term. Even though online evaluations allow students to respond at their convenience, response rates continue to decrease (Avery et al., 2006; McGourty et al., 2002a, 2002b; Sax et al., 2003; Thorpe, 2002). As response rates continue to decline (Dillman et al., 2002), data on nonrespondents can be just as important as the data collected about respondents (Berinsky, 2008).

Survey Error

Error attributed to nonresponse is one of six possible errors in the life cycle of a survey, all of which can affect the statistics, that is the "the quantitative summaries of observations" (Groves et al., 2004, p.2). Figure 2.1 illustrates the two major parts of the survey process that produce the resulting statistic: measurement and representation. Survey research is susceptible to a form of error in every step as demonstrated in this

diagram, from the beginning at the top when a researcher decides who and what will be the focus, ending with the production of the survey statistic at the bottom. The highlighted items are errors most associated with the SET process.



Adapted from: 1) *Survey Methodology* by R.M. Groves, F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau, (2004), p. 48; and 2) Getting back to the basics of survey research, in P.D. Umbach's *Survey Research: Emerging Issues* (2004), p. 92.

Figure 2.1. **The survey process: Potential errors.**

The left side of the figure, the measurement aspect, consists of four steps that are part of a survey's life, and Groves et al. associate three errors with this section. The first step is construct, which refers to the topic of the survey; it is what the researcher wants to know. In the second step, the study then must decide how to gather that information or measure the construct. Validity, a major concern with survey developers, is the issue signaled between the first two steps. Threats to the validity can occur when the measures do not gather information to appropriately research the topic of the study. Measurement error occurs with data collection when the information gathered does not accurately report real or true values. The fourth step, edited response, is a product consisting of the data used to perform analysis and develop survey statistics. However, processing errors can occur before this product is finalized. Data analysts make certain judgments to omit cases or change answers to missing data as they check data for outliers and other responses that may be inaccurate, and they may process this information incorrectly leading to this kind of error.

Representation (on the right side of the diagram) consists of five parts of the survey process that produce the survey statistic. First, by choosing the target population, a research selects the set of persons meriting study. The sampling frame, selected from the target population, contains the units or groups who could participate. Third, a sample is selected from the frame to participate, perhaps by answering questions in an interview or by completing a survey. Those who choose to participate become the respondents, providing a clearer picture of these individuals in relationship to the topic of the study. Before moving to the last consideration in the representation perspective, it is important

to note that not all surveys select participants in this same way. While many target populations (such as homeowners or working adults) require sampling procedures, some target populations are small enough to survey without necessitating a sample selection. For example, in this study, the institution sends SETs to all students.

There are two errors associated with choosing participants: coverage and sampling errors. Sampling error, according to Groves et al. (2004), is deliberate when all persons in the sampling frame cannot be measured. However, there may be groups that lack representation, and sampling error becomes a larger concern. The SET process typically includes all course participants, and sampling is rare. One of the two errors more common with SETs is between the target population and the chosen sampling frame, coverage error. As previously discussed, coverage error was more prevalent in paper surveys: undercoverage happened when not all students were present on evaluation day and overcoverage happened when individuals outside of the target population would complete surveys or when class members would complete more than one evaluation. Nonresponse error is the second major concern with SETs. Now often associated with web-based evaluations, nonresponse error occurs when results from respondents differ from the results obtained from the sample. Bias associated with nonresponse occurs when this error is systematic across particular groups.

To compensate for sampling, coverage, and nonresponse errors in the representation and as a last step in this lifecycle, researchers conduct post survey adjustments. To improve the sample estimate and survey statistic, utilizing methods such as weighting responses from underrepresented groups and adjusting for nonresponse is

common. Adjustment error is the final error addressed in this diagram. Incorrect or inappropriate adjustments are possible when true information about the target population remains unknown. Data analysts may delete missing observations or outliers or impute for missing data, but error occurs when these methods are not exact.

Each step in the measurement and representation areas throughout the survey process is prone to a type of error. With student evaluations of teaching, the paper-based mode of collection saw coverage as the most common error because not all students had a chance to respond. Replacing this more expensive and time-consuming mode with an online system significantly reduced coverage error. Web-based surveys to postsecondary students allow for most, if not all, students to respond. However, this change has introduced an error that had previously not been a large concern: nonresponse. With these issues in the forefront, the focus of this study remains on nonresponse.

Nonresponse Error.

Survey nonresponse, and the potential for error therein, is on the rise, and it is increasingly receiving the most attention from survey researchers (Dillman et al., 2002; Groves et al., 2004; Groves & Peytcheva, 2008). From the diagram, nonresponse error is a representation error. If 100% of the sample does not participate in the survey, nonresponse error is possible, with increasing associations with web-based evaluations (Avery et al., 2006; Kucsera & Zimmaro, 2008; Vehovar et al., 2002). If nonrespondents have the potential to answer questions differently than respondents, nonresponse error is present.

Bias is the most common form of nonresponse error, occurring when nonresponse is systematic across particular groups; that is, it occurs when nonrespondents are different from respondents (Groves, 1989; Groves et al., 2002; Groves & Peytcheva, 2008; Porter, 2004b; Rubin & Zanutto, 2002; Umbach, 2004, 2005). Survey research reveals significant biases in how they responded as well as in response rates among types of people, types of institution, and other characteristics (Goyder et al., 2002; Moore & Tarnai, 2002; Porter & Whitcomb, 2005). The following sections on nonresponse influences, theories, and studies more thoroughly attend to bias.

Influences on Nonresponse

Most surveys can yield two different types of nonresponse: unit and item. Unit nonresponse occurs when a member of the target population does not submit the survey (Groves et al., 2004). The second type of response can occur if a participant submits the survey with only a portion of it completed. Item nonresponse refers to skipped questions or even items answered, "I don't know" or "no opinion" (Berinsky, 2008). The focus of this study was unit nonresponse, to predict the likelihood of submission of online SETs. At NC State, unit nonresponse has been the main concern to campus administrators and researchers.

The decision to respond to a survey request relies on a variety of factors dependent on each person solicited for response. Two sets of influences on the decision to respond come from Beatty and Herrmann (2002) and Groves and Couper (1998). According to the former, there are three broader influences on potential participants enabling or hindering them from completion. Most of these are item-response specific,

usually referring to individual questions on the survey instrument (Beatty & Herrmann, 2002). The first, the influence of cognition, addresses the ability of the person to respond or even respond correctly (see also Schwarz, 2008; Tourangeau, Rips, & Rasinski, 2000). This includes access, understandability, and availability of the information to complete the survey. Second, the adequacy judgment factor may affect people who cannot be certain about an answer, and they must guess. The third factor is communicative intent. Some respondents may not be as detailed as desired, possibly due to lack of time, lack of ability to express themselves, or reluctance to tell the truth.

The second set of influences appeals more to unit nonresponse than the previous set, especially as is particular to this study. A more holistic approach to understand influences of cooperation is the four blocks (Dillman et al., 2002; Groves & Couper, 1998). The first two of four blocks are variable within the survey sample; that is, these influences on participants may determine the level of response. First, social environment influences may prevent response to surveys. Societal attitudes, beliefs, cultural norms, and demographics may all have an impact on the decision to respond to a survey. For example, Groves and Couper (1998) report that small town residents are more likely to cooperate than those in an urban setting, and those in rural locations will probably respond at a higher rate than any. Second, persons have unique characteristics, knowledge, and attributes influencing their participation. For example, individuals with a higher sense of civic duty or an interest in the survey's topic are more likely to respond (Dillman et al., 2002).

The third and fourth blocks of influence are more dependent on the survey design and interviewer. Mode of collection is a third factor in determining participation. Face-to-face surveys may yield higher response rates than online surveys (Avery et al., 2006; Kucsera & Zimmaro, 2008; Vehovar et al., 2002). Also, higher responses will result from reminders to complete the survey at various intervals that the survey is available (Berinksy, 2008). Finally, research demonstrates that, in face-to-face interviews, an interviewer effect is present. The interviewers that establish a rapport with the interviewee "are more likely to secure interviews" (Berinsky, 2008, p. 311; see also Groves et al., 2002).

Theories of Nonresponse

Patterns of influences and previous research have shaped several theories of survey participation, particularly why some people respond and why some do not. Well-established theories allow analysts and scholars to understand their findings better and continue to inform research. For example, the leverage-saliency theory of participation describes the factors influencing individual participants in their decision to respond to a survey (Groves et al., 2000). Also, academic disciplines (Smart et al., 2000), may explain nonresponse by examining environmental influences associated with student majors. Survey fatigue may also explain nonresponse in the modern context, as survey requests continue to rise with online instruments (Groves et al., 2004; Porter et al., 2004). But first, below are five theories that Dillman et al. (2002) discuss as most important to consider.

Opportunity Costs

When a potential respondent is in the process of making the decision to complete the survey, they are probably using "opportunity costs" to make the decision. Lack of time and higher priorities (other than taking a survey) are part of this. At the end of a semester, multiple papers, projects, and tests are due and students are packing to go home over break. At the same time, the university emails students wanting them to complete a survey for each course they are taking. Students may choose their other responsibilities such as a job, coursework, or an end-of-semester get-together with friends instead of responding to the surveys. For example, those with jobs may find it more difficult to complete surveys due to time constraints (Couper, Kapteyn, Schlonlau, & Winter, 2007).

Social Exchange

Social exchange theory, the second concept, originally described behavior in the fields of sociology, anthropology, and psychology (Emerson, 1976). Blau (1964) viewed all interactions between people as either giving or taking, and that such actions seek reciprocation. Dillman (1978) saw social exchange theory as beneficial in explaining nonresponse (see also Dillman et al., 2002, 2009; Groves, Cialdini, & Couper, 1992). Choosing to respond may occur due to feelings of duty or wanting to help someone or the institution. For example, college teachers may convey the idea that SETs mean a lot to them and that their students' opinions truly matter to them. Students may also see their friends complete SETs or participate in related discussions and choose to respond due to social pressures and attitudes. Porter and Umbach (2006a) applied the social exchange theory to their study of survey response rates by institution. They found that certain types

of institutions, such as women's colleges and selective institutions, had higher response rates. Porter and Whitcomb (2007) found that participants with a strong relationship to the institution were more likely to respond than those with a weaker relationship. The sense of duty students felt towards their institutions may have increased response rates; in turn they reciprocated services rendered by the institution (Blau, 1964). At NC State, students in social or humanities disciplines may have responded more due to this natural civic sense of duty. Or, perhaps students with greater ties to campus, such as living on-campus or being a student-athlete, may have responded at a higher rate because they see the need to give back to their institution.

Personal culture, not just institutional culture, may also explain nonresponse. Patterns of social participation such as minority group oppression, social distance perceptions, and helping behaviors have explained nonresponse survey research (Johnson et al., 2002). Minority group oppression represents feelings of mistrust and resentment towards a majority group. Barriers to the advancement of these groups "such as poor education, poverty, and discrimination" may also translate to barriers in the survey process (p.66). Social distance refers to differences in social interactions influenced by perceptions, attitudes, beliefs, status, and demographics. For example, mixed-gender interviews are not culturally appropriate in many societies, and cultural cues may be missed if the interviewer is ignorant of certain traditions and customs. In web-based surveys, this may not be a large concern, though survey design may overcome some concerns related to social distance. Lastly, helping behaviors exemplify more community-based attitudes; in survey research, people with these behaviors are more

likely to participate or respond. Urban locations are less likely to exhibit these helping behaviors (Groves & Couper, 1998; Porter & Umbach, 2006a). For the purposes of this study, attributes such as race, nationality, and whether the student is an international student will be available and may show evidence of nonresponse patterns related to SETs.

Salience

The third concept involves the salience of the topic and survey. Overall, research has demonstrated that response is more likely if the topic of the survey or questionnaire is salient to the potential respondent (Groves et al., 2006; Groves & Peytcheva, 2008; Groves, Presser, & Dipko, 2004; Heberlein & Baumgartner, 1978; Remedios & Lieberman, 2008). In university SETs, each survey has the same basic set of questions. However, examples of salience in SETs may be examining the student's personality type or comparing the course location with student's major. A philosophy major may be more inclined to complete the end-of-course evaluation for a philosophy course, but the same philosophy student may fail to respond to the same evaluation for a science class. There may be predispositions in students that increase the likelihood of response, measured by academic environment of their major or personality types (Porter & Whitcomb, 2005; Smart et al., 2000).

Interviewer Effects

Fourth, interviewer effects may influence response rates, since cooperation rates can vary for interviewers (Dillman et al., 2002; Morton-Williams, 1993). An interviewer (for face-to-face or telephone surveys) and the survey design (for paper or web-based surveys) can have a significant influence on the participant to complete the survey

(Groves et al., 2006). Perhaps shared characteristics between the interviewer and the interviewee (age, language, race) influence participants to be more open. Additionally, the demeanor, friendliness, and established rapport help a person determine if they will participate, and then if they will fully and freely participate (Groves & Couper, 1998). Morton-Williams (1993) found that tailoring introductions to potential respondents, strong social skills, and maintaining contact were important characteristics and steps that interviewers could take to influence response rates favorably (see also Dillman et al., 2009). Interviewer effects appear to pertain less to web-based surveys, but there are several connections possible. A welcome message, friendly directions, and willingness to answer questions or help with an online survey may increase the likelihood of response.

Sequential Design and Nonresponse

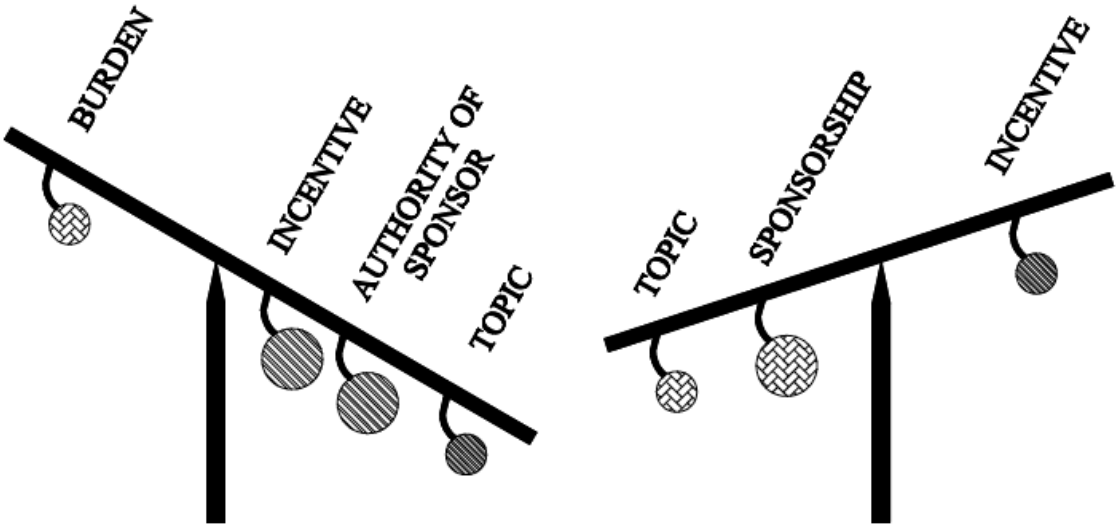
Lastly, sequential survey design is the fifth and final concept in nonresponse theories. Longitudinal and establishment surveys administer follow-up surveys to the same participants, typically over a period of time. If a respondent had an unsuccessful experience with a previous survey, or if they do not agree with the results and decisions made from previous surveys, they may choose to discontinue their participation (Dillman et al., 2002). Though SETs occur at the end of every term, they are not sequential and counted as part of one study.

When studying nonresponse in SETs, opportunity costs, social exchange, and survey salience are the most applicable types of theory from this section. In this particular study, opportunity costs are difficult to measure since there will be no contact with nonrespondents. Variables not included in the study are whether students have jobs,

dependents, or other time constraints that hinder their ability to participate in the course evaluation. This is an area for potential follow-up with nonrespondents, but would be beyond the scope of this study. Social exchange theory, academic environment, and survey salience were helpful in framing the outcomes of this study.

Leverage-Saliency

First, the theory of leverage-saliency (Groves et al., 2000) explains that a person's decision to cooperate is based on multiple aspects unique to each individual. The survey design, a person's situation, and their characteristics may give an individual the leverage to make the decision to participate in the survey, but someone else may decide not to take that same survey based on the same aspects. This theory represents a holistic, contextual, and individualized explanation of survey participation. Several variables, characteristics, and attributes may influence different people in different ways as exemplified in Figure 2.2 below. In this study, those influences may extend beyond individuals to certain populations, types of people, and demographics.



Adapted from: 1) *Survey Methodology* by R.M. Groves, F.J. Fowler, M.P. Couper, J.M. Lepkowski, E. Singer, and R. Tourangeau, (2004), p. 177; and 2) Leverage-saliency theory of survey participation by R.M. Groves, E. Singer, & A. Corning (2000).

Figure 2.2. **Leverage-saliency: Participation for two individuals.**

Figure 2.2 demonstrates what may influence two individuals to complete the same survey (Groves et al., 2004). One scale represents each person. A scale tilting to the right indicates survey participation, and the items on the right side of each scale are factors in deciding to participate. Tilting to the left indicates refusal, and factors on the left side of the scale are influences of incompletion. Ultimately, the tilting of the scale indicates the final decision, but the other factors taken into consideration, whether or not they aligned with the ultimate decision, are included in the diagram indicating what matters to each individual. In the figure, Person 1 decided to participate in the survey because their scale tilts to the right, but Person 2 did not participate as their scale tilts to the left. This figure

shows that two survey aspects, topic and the survey sponsor's authority, influenced Person 1 to participate. However, the same two aspects for the exact same survey caused Person 2 to decline participation. In addition, the survey was a burden to Person 1, but prioritization of the other three characteristics caused him or her to participate. Burden was not even a factor in Person 2's decision to decline. This theory and its depiction demonstrate that each individual is unique. The influences on their unique decision to participate or decline depend on circumstances related to each individual, attributes of survey design, the survey topic, and personal preferences. In studies of SET nonresponse, it is possible that attributes of SETs influencing individuals to respond also influence particular populations to respond.

While this figure mostly focuses on survey attributes, personal and contextual factors may also influence the survey participation decision, such as those addressed in opportunity costs, survey fatigue, and social exchange. In the current study, these influences may include the instructor and the burden on the student. All of these factors influence individuals differently, if at all. Perhaps certain patterns of characteristics illuminate the choice to respond to or ignore evaluations of teaching, revealing bias among groups. Those potential biases as applicable to NC State emerged in multilevel models; chapters four and five further discuss these outcomes.

Academic Environment

Second, Holland's theory of careers as applied to academic majors and disciplines may also help explain nonresponse at postsecondary institutions. Originally, Holland's work in vocational choice demonstrated that peoples' personalities best matched with

certain careers (1959, 1966, 1973, 1997). He characterized people based on six types: realistic, investigative, artistic, social, enterprising, and conventional. Realistic types avoid interpersonal interactions and prefer activities with predictable outputs; they typically choose careers that are more hands-on, mechanical, or practical. Investigative people seek to understand and control their surroundings and avoid social activities, and their careers are typically analytic or scientific in nature. Artistic people do not prefer order and routine, but instead look for ways to express themselves creatively. Painters, dancers, sculptors, and other creative disciplines are represented in this career field type. The social personality type avoids ordered activities with predictable outputs, and these people have a higher capacity for human interaction. Associated career fields include education, the humanities, and others associated with helpful or nurturing traits. Enterprising types of people are extraverted and very politically oriented. They manipulate to attain their goals, are unscientific, and often choose careers in business, law, or journalism. Lastly, conventional people enjoy structure and place significant priority on financial achievements and power. They can also choose careers in business, but may focus on finances, accounting, or secretarial work.

Academic disciplines can characterize students' personality types according to majors using the same six classifications: realistic, investigative, artistic, social, enterprising, and conventional. In *Academic Disciplines* (2000), Smart, Feldman, and Ethington examined academic fields of study based on Holland's six types (see also Rosen et al., 1997; Smart et al., 1999). By avoiding incompatible situations, students tend to pursue major environments that appreciate and align with their skills and attitudes. In

return, faculty members nurture inductees by upholding the discipline's norms and values; they acculturate and socialize students while reinforcing common language and beliefs. Attributing Holland's six types to academic disciplines is a way to examine student behavior, as in this study, survey and SET participation. For example, investigative students may respond at a higher rate. Scientific disciplines associated with this type are empirical by nature and would perhaps see the value in contributing to research. Design majors may not see the importance of investigation and scientific query and may decline to take the survey. Submission of SETs may be more prevalent for some types than for others.

Several additional studies of postsecondary students have demonstrated that personality influences behaviors and decisions (Astin, 1993; Holland, 1997; Porter & Umbach, 2006b; Smart et al., 2000; Umbach & Milem, 2004; Yu et al., 2007), including survey participation (Porter & Whitcomb, 2005; Sax et al., 2003; Umbach & Porter, 2002). As examples, English, Fine Arts, and Humanities students had higher response rates in one study (Sax et al., 2003), and extroverts were more likely to respond in another (Marcus & Schutz, 2005). Porter and Whitcomb (2005) framed nonresponse around the theory of careers and academic disciplines, as this study does. They found that the decision to respond to the National Survey of Student Engagement (NSSE) correlates to those six personality types, as associated with their major academic field of study. Majors within the artistic fields were less likely to participate in the surveys associated with their study, and students within investigate majors were more inclined to participate. The current study uses the same typology as Porter and Whitcomb; that is, the

personalities developed by Holland. The propensity to respond as associated with personality type (inferred by major) was significant.

Survey Fatigue

Also called "over-surveying", survey disillusionment, or survey saturation, survey fatigue is the third possible explanation for survey nonresponse (Groves et al., 2004; Porter et al., 2004). Surveys today seem omnipresent; we receive them while browsing the internet, in businesses we visit, at the end of a shopping experience, and for services we receive. Such saturation may indicate that our society consequently is succumbing to survey fatigue (Groves et al., 2004, 2009; Porter et al., 2004). Moreover, survey creation by untrained methodologists can effect poor design, which may contribute to increasing declines in survey participation. Surveys that immediately succeed others may be most susceptible to low response rates (Porter et al., 2004), however the number of surveys actually requested is difficult to calculate since campus and non-campus related surveys may reach students' email inboxes every day. Students may be most susceptible to survey fatigue on postsecondary campuses where the student is viewed as a consumer, quality assurance is important, federal and state measures assess students, SETs are collected, and national surveys take place.

In this study, several measures of survey fatigue are available, including number of credit hours, number of SETs to complete for that semester, and student or class (e.g. sophomore, senior). If students received an abundance of evaluations to complete at the end of the semester, they may feel overwhelmed or fatigued by this process. Students with a higher number of credit hours or who have been at the institution longer may feel

they have completed “enough” SETs, consequently resulting in an unsubmitted survey (Porter et al., 2004). At NC State, perhaps some students succumbed to fatigue and only completed one or two SETs even though five evaluations were available to them. These partial participants may have felt that they completed as much as needed (Laurie, Smith, & Scott, 1999; Porter et al., 2004).

Using a web-based instrument has increased the ease of contact and delivery, but this mode has also increased the number of survey requests that students receive. Resulting feelings of ambivalence towards surveys are indicative of survey fatigue, which may influence nonresponse in SETs and other surveys. Survey fatigue is a newer area in nonresponse research, especially in SETs. This may be due to the surveys’ increased presence and ease of delivery in online environments and the difficulty that survey researchers have in measuring all survey requests administered. Other surveys of postsecondary students demonstrate the potential for survey fatigue (such as Porter et al., 2004), but have excluded SETs in their research. There is a great need to increase research related to survey fatigue and other emerging issues in survey nonresponse of postsecondary students, especially SETs and online surveys. Below are the few that have addressed SETs at colleges and universities.

Studies on Survey Participation at Post-Secondary Institutions

Survey participation and survey ratings, that is, if they respond and how they respond, seem to share some findings. For example, high achievers tend to rate more favorably their instructors (Cohen, 1981; Costin, 1978; Crumbley & Reichelt, 2009; Isely & Singh, 2005; Marsh, 2007; McKeachie 1969), and they are more likely to participate,

or submit their SET, at all (Avery et al., 2006; Fidelman, 2007; McGourty et al., 2002a, 2002b; Porter & Umbach, 2006a; Porter & Whitcomb, 2005). Females are more likely to respond than males (Avery et al., 2006; Fidelman, 2007; McGourty et al., 2002a, 2002b; Porter, 2004b; Porter & Umbach, 2006a; Porter & Whitcomb, 2005). Lastly, white students generally respond at higher rates than students of color (Avery et al., 2006; Fidelman, 2007; Porter & Umbach, 2006a).

Other factors have been influential in college students' decisions to submit surveys. Students' personality types and majors have displayed different propensities to respond (Marcus & Schutz, 2005; Porter & Whitcomb, 2005; Sax et al., 2003; Yu et al., 2007). Institutional characteristics have also made a difference in students' decisions to complete SETs (McGourty et al., 2002a, 2002b; Porter & Umbach, 2006a). Some colleges have greater technology demonstrations, award incentives, and better communicate to students that their input is needed (McGourty et al., 2002a, 2002b; Thorpe, 2002). Other institutional factors linked to lower response rates include an urban location, public status, and an increased population of students per acre (Porter & Umbach, 2006a).

However, there remains an evident lack of literature that has focused on factors of nonresponse in web-based SETs. There have been two recent dissertations on nonresponse to SETs. While both studies used multilevel modeling techniques to look at nonresponse, they were very different from the current study. The first dissertation incorporated nonresponse into only half of the study; predictors of students' ratings comprised the other half (Jones, 2009). The researcher chose a sample of about 2000

students based on class size, class time, and credit hours. She used the university's course evaluation instrument, but only incorporated nine variables to determine nonresponse. Several predictors of nonresponse emerged: gender, ethnicity, and final course grade.

The second dissertation examined both undergraduate and graduate students at Boston College (Fidelman, 2007). Again, the focus included students' ratings, but also compared the two modes of data collection: paper and web-based. Twelve classes (over 800 students) were randomly selected if they met two criteria: 1) had more than two students and 2) were three or more credit hours. The online course evaluation period lasted a week, after which students completed the same course evaluation on paper in class the next week along with a course evaluation attitude questionnaire. Results found that gender, expected grade, year in school, and teaching experience were predictors of nonresponse.

Fortunately this study incorporated more variables, more students, and more course evaluations, effecting more generalizeable results. Moreover, these studies took place at private institutions and limited their sample size to certain classes and majors. Focusing only on nonresponse allowed for expanded results, a larger dataset, and clear outcomes. The institution and the SET process for this study undergo further examination below.

ClassEval

NC State, the setting for this study, has developed an online system for SETs called ClassEval. The University Planning and Analysis (UPA) office at NC State

acknowledges three purposes of ClassEval results (2008d). First, administrators and committees consider the aggregate results of SETs for curricular and personnel decisions. Second (and most common for universities and colleges), faculty members receive their class results and can make adjustments to improve their teaching in the future. Third, faculty can also use SETs to apply for certain grants or awards. The risk for nonresponse error, potentially resulting in poor data quality, increases as response rates decrease. Therefore, administrators and faculty desiring to improve their teaching, their courses, and the quality of education offered, necessitate accurate and valid results for decision-making processes.

ClassEval Process

NC State has employed SETs since the fall of 2000 with a common set of core questions (UPA, 2008a). In paper form from 2000-2006, colleges tended to house this process. Students completed these evaluations using class time, and response rates ranged anywhere between 5-200% (UPA, 2008c). Data quality proved to be very poor, and 15% of the data was unusable. In the spring of 2007, the university moved to an online delivery method called ClassEval, which utilized open and closed questions, unique questions (tailored to each individual instructor), and core questions. Closing at 8am on the day of final exams, the survey period typically lasts 2-3 weeks, and students have to log in with their NC State user name and password. According to a UPA report (UPA, 2008b), around 40-45% of students did not respond to the ClassEval survey during each period of three semesters from Spring 2007 to Spring 2008. For the Fall 2009 semester, UPA administered 134,929 SETs to students. The lowest response rate since the online

implementation of ClassEval occurred during this semester, with only 48.9% of the surveys submitted.

Research on ClassEval

NC State's UPA office has answered a few basic research questions pertaining to ClassEval in a presentation of their results on their website (2008c). In 2005 and 2006, the office discovered that there was no statistical difference to the response rates on paper versus online, though the online response rate was higher overall by about four percentage points. However, these results are incomparable as discussed above with 5-2000% response rates. Secondly, UPA found that if response rates for a class exceed 50%, nonrespondents would not vary from the average ratings. That is, the scores stabilize even after adding in more respondents after 50% have already responded, they are representative of how the entire class would respond, and are therefore meaningful.

In the same presentation (2008c), UPA reported that on-campus classes, College of Management courses, and students with higher grades have much higher response rates. On the other hand, distance education classes, laboratory classes, small classes, lower level classes (100 and 200 level courses), and students in the College of Design, the College of Textiles, Music and Physical Education all had significantly lower response rates.

Summary

These results require replication and further analyses. Online SETs have their many advantages, but they also expose the following concerns about student nonresponse:

1. Research suggests that, in general, online surveys obtain lower response rates when compared to paper administration. Lower response rates increase the risk for nonresponse error.
2. A variety of factors influences the decision to complete surveys, and theories related to priorities, over-surveying, culture, and academic environment are available to further inform studies of postsecondary survey nonresponse.
3. Some nonresponse studies conducted with college populations have reported statistically significant differences in participation by student gender, GPA, and ethnicity. Little research is available about nonresponse on college campuses, especially in online SETs.
4. There is a critical need to conduct campus-based research about institutional nonresponse, especially considering the recent migration to online instruments and the potential high stakes of course evaluation results.

As our society's saturation in the online environment increases and as survey requests increase, what do nonrespondents at NC State look like? What factors contribute to a submitted or unsubmitted SET? Current research needs to examine influences of nonresponse, as much has changed in recent years.

Considerations for Research

With the relative newness of web-based surveys, the decline in survey response rates, and the effects of the survey results, it is evident that there are opportunities to add

to the research about nonrespondents. First, recent nonresponse studies on web-based SETs have few variables with which to work (Fidelman, 2007; Jones, 2009). The addition of more variables to statistical models may increase our knowledge of nonresponse. In this study, variables such as environment of the major, athletic membership, housing, and others were available for analysis. Second, many of the previous studies only focused on a few classes at the participating college or university (Dommeyer et al., 2004; McGourty et al., 2002a; Thorpe, 2002). Institutional researchers must increase their knowledge about the whole population, not just a few sections or classes. The current study took place campus-wide, with data from all undergraduates. This difference in data size also allows for the opportunity to explore whether nonresponse and nonresponse bias patterns are similar across the student populations at this university.

This Study

The focus of this study was to address these issues and examine factors related to unit nonresponse in online SETs. No comprehensive studies of SETs at large, public universities have focused on all undergraduate nonrespondents. This dissertation study examined thousands of SETs at NC State for patterns in nonresponse. In the next chapter, I discuss the steps to obtain and analyze the data in exploration of nonresponse.

CHAPTER THREE: METHODOLOGY

Survey participation is on a steady decline (Dey, 1997; Dillman et al., 2002; Groves et al., 2004, 2009), and surveys administered to postsecondary students are no exception (Avery et al., 2006; Porter, 2004b; Sax et al., 2003). In the age of accountability and data-driven decision-making processes, error resulting from nonresponse can severely affect the data quality and lead to incorrect conclusions. Results from student evaluations of teaching (SETs) are components of personnel records, promotion and tenure decisions, and are integral parts of the process through which instructors go to improve their instructional methods. In an effort to better understand SET participation at North Carolina State University (NC State), this study examined several potential factors of nonresponse. If survey administrators better understand the population to whom they administer the SETs, they may know how to reach them to increase response rates (Nulty, 2008).

This chapter describes the manner and processes that constituted the study and data analysis. The purpose of this study was to investigate unit nonresponse for end-of-semester SETs by exploring factors that contributed to the probability of response. Guided by the following research questions, I analyzed the data using a variety of statistical tests and methods:

1. To what extent was the likelihood of response to SETs related to student-level measures, such as demographics (e.g. age, gender, race) or academic performance (e.g. GPA, SAT score, grade in course)?
2. To what extent was the likelihood of response related to personality and academic environment (Holland, 1973, 1997; Smart et al., 2000)?
3. To what extent did salience (e.g. if course is in the same department as the student's major, student grade in course) increase the likelihood of response? Did the relationship vary significantly between students? Could student-level measures explain this variation?
4. To what extent did survey fatigue (e.g. number of SETs to complete, class rank) predict response rates? Did response rates correlate to opportunity costs (e.g. number of credit hours, number of courses taken that semester)? To what extent did culture (e.g. environment, institutional engagement, personal background) influence response rates?

Much is at stake when considering SET submission in this context. Administrators make personnel and academic decisions with the results in consideration, and instructors incorporate results to improve teaching. Improving response rates decreases the risk of nonresponse error and tends to improve data quality. To appropriately address nonresponse on campus, institutions must first understand its factors and influences.

Study Design Overview

Institution, Participants, and Administration of Survey

NC State is the largest university in the state of North Carolina with over 32,000 undergraduate and graduate students. Founded in 1887 as a land-grant college, it is located in downtown in the capital city of Raleigh in the Research Triangle region of the state. At NC State, 10 colleges house 115 undergraduate degree programs, 162 master's degree programs, and 62 doctoral degree programs, several of which are nationally ranked. The institution boasts a strong commitment to the land-grant mission, with many research, technology, and extension programs.

NC State's University Planning and Analysis (UPA) office is the institutional research office charged with administering class evaluations since the fall of 2000 (UPA, 2008a). At this time, the Evaluation of Teaching Committee (EOTC), a standing university-wide committee, recommended completely centralized evaluations, but some of their recommendations were cost-prohibitive. In paper form from 2000-2006, students completed SETs using class time, and response rates ranged anywhere between 5-200%, with the average at 46% (UPA, 2008c). Moreover, many colleges or departments did not follow the standardized, core questions recommended by the EOTC. About 15% of the data collected was of low quality, therefore unusable and incomparable. These issues along with inconsistencies in administering the surveys, unsecure procedures, and unsecure collection methods, contributed to the need for a better course evaluation system.

To address these issues, the College of Agriculture and Life Sciences (CALS) developed an online system. Other colleges and departments piloted this online instrument in the Fall 2006 semester, and by the following semester it was available to all departments and colleges. In the spring of 2007, the entire university moved to this online mode and called the survey instrument ClassEval (Appendix A). The web-based mode was a more cost-effective and efficient strategy allowing for centralization of the collection process. The ClassEval instrument, still in use for this study, utilizes open-ended and closed-ended questions. Core questions are standard, but instructors have the option to add unique questions.

Data Collection.

For this study of the Fall 2009 semester, the UPA office began data collection two weeks before the final exam period started. They sent an initial request (Appendix B) to students via email from the email address classeval@ncsu.edu. The email contained a link that, when clicked, brought students to a web page where they had to enter their university username and password to access the ClassEval surveys. A list of surveys available for them to complete appeared on the screen. As students submitted each course evaluation, the list updated their progress and showed the students which surveys remained incomplete. From the same ClassEval email address, the UPA office periodically sent reminder emails to students who still had SETs to complete, a common technique for survey administrators to increase response rates (Dillman et al., 2002). Appendix C lists when the reminders were emailed, and Appendix D contains the text of the reminders. Throughout the data collection period, students who had already submitted

all their evaluations did not receive these reminder emails. UPA also employed other efforts such as emailing the faculty about ClassEval (Appendix E) and non-email reminders (listed in Appendix F) in efforts to combat nonresponse. The data collection period lasted three weeks, opening at 8am on November 19 and closing at 8am on December 9, the day final exams began.

Dataset Creation.

Each semester, the UPA office prepares reports from course evaluation results for the faculty in aggregate form. For this study, UPA prepared a dataset that included about 25 independent variables based on course and student attributes. Participants in ClassEval were only undergraduate courses and undergraduate. Independent studies and other one-on-one classes are never eligible for evaluation to keep students' identities unknown.

Cases, or observations, in the dataset are evaluation-based. In other words, each case in the dataset represents one ClassEval form. If a student took five classes, each SET represented one case, totaling five cases represented by that one student. If a class contained 25 students, there were 25 observations for that class. Both students and courses had unique, untraceable identifiers to avoid direct identification of the course, student, and faculty instructor, and variables did not identify any student or faculty member. Students have expressed concerns over confidentiality in their responses to SETs (McGourty et al., 2002b), and research project review boards (i.e. IRB) require appropriate actions to protect these students. Participant confidentiality protects both records and student identification, and making this known to them likely increases response rates to surveys (Anderson et al., 2005; Groves & Couper, 1998; Groves &

Peytcheva, 2008; Nulty, 2008; Quinn, 2003). Because these variables have been limited to only de-identified student data and course data, no instructor data were available. Instructor attributes have previously demonstrated differences in how students respond (Nilson, 2003), but extended research specific to SET participation is not available. The UPA office prioritized confidentiality for all involved.

To further ensure protection of the identities of students and instructors, I submitted a research proposal to the NC State Institutional Review Board (IRB) office in March 2010. After reviewing the proposal, the IRB office determined that this study was exempt from their oversight, as long as the participants' role did not change. That is, as long as they were not contacted in this study. The appendices include this IRB proposal (Appendix G) and the exempt letter received from the IRB office (Appendix H).

Variables.

An abundance of data was available to this study because ClassEval utilizes a university-wide login for each student, linking their campus and personal information. For that reason, this study was able to incorporate more variables than any other study of nonresponse on college campuses. The main dependent variable for this study was dichotomous, based on whether the student responded to ClassEval for their class (0=no response, 1=submission of the evaluation). Over twenty-five independent variables (Table 3.1) designate characteristics of the students and the courses.

Table 3.1. **Variables linked to nonresponse theories.**

Variable	*Variable Description*	*Theory Linked to Variable*
Age	Student age at semester Census (10th day of classes)	Demographics
Gender	Gender of student	Demographics
Race	Race of student	Demographics
Semester GPA	Combined GPA for all courses taken that semester	Past Research
Transfer Student	Did student attend another post-secondary institution before NC State	Culture
Student grade in course	The grade the student received in the course, measured on a 4-point scale.	Past Research Opportunity Costs
Class Rank	Year/level of undergraduate student by hours passed (e.g. sophomore, senior, etc.)	Past Research Survey Fatigue
Total SCH	Total credit hours passed at all institutions that NC State accepts	Opportunity Costs Past Research Survey Fatigue
NC State SCH	Total credit hours passed at NC State	Opportunity Costs Past Research Survey Fatigue
Semester SCH	Credit hours taken in semester	Opportunity Costs Survey Fatigue
Major	Student's primary major on Census (10th day of classes)	Academic Disciplines Topic Salience
Entry Semester	First semester that student entered NC State at undergraduate level	Past Research
Semesters Enrolled	Number of semester attended course at NC State (excludes summer school)	Survey Fatigue
Course Location/Env.	Is the course in the student's department?	Academic Disciplines Topic Salience
Number of Courses	Number of courses taken that semester	Opportunity Costs Survey Fatigue
Athlete	Is student a scholarship or non-scholarship student-athlete	Opportunity Costs Academic Disciplines

***Table 3.1*. (continued)**

Variable	*Variable Description*	*Theory Linked to Variable*
Campus Housing	Is student in campus housing? Or commuter?	Social Exchange
TOEFL *	Did student take English as a foreign language exam?	Culture
SAT	Highest SAT score combination	Past Research
Nationality	Student's nationality	Demographics Culture
Residency	NC resident for tuition purposes	Culture Social Exchange

* Only three students took this exam, eliminating this variable from analysis.

Linking Variables to Theories: A Conceptual Approach to Data Analysis

Literature and theories support the variables included in a dataset. Variables chosen for this study were consistent with established theories on survey nonresponse, past research on survey nonresponse, and past research on student evaluations of teaching. In the following sections, hypotheses systematically introduce and describe the variables, describe how they informed the study, tie the methodology to the research questions, and demonstrate links to theories of survey nonresponse.

Hypothesis One

Females, white students, and older students will be more likely to respond to SETs than their peers. Studies on nonresponse, in general survey research and in research on postsecondary student populations, have shown consistently that certain demographics tend to increase the propensity of individual response. Gender tends to be the strongest,

as females usually respond more than males (Dey, 1997; Johnson et al., 2002; Jones, 2009; Lepkowski & Couper, 2002; Moore & Tarnai, 2002; Porter & Whitcomb, 2005; Sax et al., 2003). In a recent study on course evaluation nonresponse, Fidelman (2007) found that females were twice as likely to complete their online evaluations. Race and nationality are also demographic variables available. In many cases, students of color have responded less frequently than their white or Caucasian classmates (Avery et al., 2006; Fidelman, 2007; Porter & Umbach, 2006a). In Sax et al.'s study (2003), web-based surveys submissions were lower for Black/African American and American Indian students when compared to White/Caucasians, Latinos, Asian/Asian Americans, and students in the "Other" category.

In national and household surveys, the relationship between age and likelihood of response has varied (Berinksy, 2008; Dillman et al., 2002; Groves & Couper, 1998; Groves, Presser, et al., 2004; Rogelberg & Luong, 1998; Roose, Lievens, & Waege, 2007). While most studies on college student surveys have not addressed age, this study included this variable. Studies that have incorporated age have found that non-traditional students, 24-25 years of age or older, are typically more likely to respond than their younger peers (Kaplowitz, Hadlock, & Levine, 2004; Sax et al., 2008; Yu et al., 2007).

Residency, that is if the student is a resident of the state of North Carolina, is another demographic variable underrepresented in the literature on postsecondary survey nonresponse. Sax et al. (2003) discovered that attending school far from home increased the likelihood of survey response.

Hypothesis Two

High academic performance will increase the likelihood of SET response. Several studies on college student nonresponse have indicated that students with higher grades are more likely to complete surveys (Avery et al., 2006; Clarksberg et al., 2008; Dey, 1997; Fidelman, 2007; Porter & Umbach, 2006a; Sax et al., 2003, 2008). This study utilized three variables to measure student academic performance. The first was semester grade point average (GPA), the average of the student's grades they received in the Fall 2009 semester reported in GPA format. Second, the individual student grade in the course measures student achievement in the course for the administered evaluation. Lastly, the student's highest score reported to the university on their SAT reasoning test measures academic performance. High academic performance may also have a relationship with topic salience; if students performed well in a course they may feel that the evaluation associated with that course has value and merits completion.

Hypothesis Three

Opportunity costs and survey fatigue will decrease the probability of response. These two theories of nonresponse indicate that individuals ignore or de-prioritize survey participation when they lack the time to do so or when they have received too many survey requests (Dillman et al., 2002; Groves et al., 2004; Porter et al., 2004). In this study, five variables operationalized these theories: class rank (e.g. senior, sophomore), the total number of credit hours the student had earned, the number of credit hours the student had earned at NC State, and the number of credit hours each student took for the Fall 2009 semester. Though not directly tied to these theories, these and similar variables

have tended to influence the likelihood of response in other studies (Fidelman, 2007; McGourty et al., 2002b; Porter & Umbach, 2006a; Porter & Whitcomb, 2005).

The theory of opportunity costs means that potential respondents will weigh their options to determine if they can prioritize the decision to take a survey (Dillman et al., 2002; Groves et al., 2004). At the end of a semester, multiple papers, projects, and tests are due and students are packing to go home over break. At the same time, the university emails students requesting that they complete a survey for each course they are taking. Some courses with more than one instructor or with associated labs will have multiple evaluations associated with them. A student taking several classes may prioritize other end-of-semester tasks over SETs. Students who live on campus may have been reminded more from posters or the student newspaper, and they would have been given more opportunities to complete their SETs. On-campus students also have access to wireless internet across campus that off-campus students may or may not have available to them. Opportunity costs can also relate to non-academic measures of activity, represented by part-time status and athletic membership in this study.

Survey fatigue, a more recent occurrence due to the increasing accessibility of audiences via the Internet, may also result from academic activity (Groves & Couper, 1998; Groves et al., 2004; Porter et al., 2004). Surveys that immediately succeed others may be most susceptible to low response rates (Groves et al., 2004; Porter et al., 2004). Nonresponse may result from over-surveying because of apathetic attitudes and an increased burden to participate. If a student has to complete seven or eight SETs in one semester, they may feel over-surveyed. The same feeling might exist for upper-level

students, those with a higher amount of credit hours, or those who have been at the institution longer. As they matriculate through their programs, students may succumb to fatigue by completing none or some of their class evaluations. Upper classmen and partial participants may feel they have completed as much as they felt was needed (Laurie et al., 1999; Porter et al., 2004).

Hypothesis Four

The probability of response will increase for students with stronger cultural and communal ties to the campus. Individuals with strong ties to a group (e.g. society, community, culture, subculture, academic field, etc.) tend to feel invested in their group's interests and its betterment. The theory of social exchange describes participation as a social activity (Blau, 1964; Dillman et al., 2002; Emerson, 1976; Groves et al., 1992). Past studies have reported that participants with a stronger relationship to the institution were more likely to respond than those with a weaker relationship (Leslie, 1972; Porter & Whitcomb, 2007). In this study, the connection or commitment participants felt towards their institution may have increased response rates; they reciprocated services rendered by the institution through the act of submitting their SET(s) (see also Clarksberg et al., 2008; Porter & Umbach, 2006a).

Academic disciplines perpetuate a culture and society unto themselves, and the student is also part of that academic environment displayed by their major (Holland, 1997; Smart et al., 2000). Two influences on survey participation are related to academic discipline. First, social exchange theory would argue that students who are taking a course in their field would feel the need to reciprocate and support the environment. They

are therefore more likely to complete an SET if the course they took was in the same department as their major, which is another variable included in this study. Students feel invested in their academic departments, feel as if they have the ability to contribute useful information, and they feel part of the social construct when providing their feedback. Second, personality as measured by academic major influences survey cooperation. Do certain disciplines nurture or produce chronic respondents? Does the academic environment encourage students to value SET completion? Porter and Whitcomb's (2005) study measured personality type by the major chosen and found a positive relationship between survey participation and two of Holland's personality types: social and investigative. Extroverts (Marcus & Schutz, 2005) and students in science, technology, and engineering majors (Yu et al., 2007) have also exhibited the potential to respond more than their peers. Appendix I contains a list of majors and the corresponding environment and personality Holland attributes to each (Rosen et al., 1997). The chart below provides an overview of the types and how many times each one occurs in a major at NC State.

Table 3.2. **Number of majors associated with academic disciplines.**

*Type**	*Occurrences at NC State*
Artistic	4
Conventional	1
Enterprising	7
Investigative	32
Realistic	11
Social	7
Transitioning Programs (Undecided)	2

* Types originally developed by Holland (1959, 1966, 1973) and further explored for postsecondary institutional environment by Smart, Feldman, and Ethington (1999, 2000).

Culture, the way of life of a group of people, permeates throughout all settings. For example, non-academic measures in this study include if the student lived on campus and if the student was an athlete for the university. These students have very strong ties to the institution, and this investment may increase their likelihood to provide social exchange in the form of completing SETs. Other patterns of social participation such as minority group oppression and helping behaviors have fielded discussion and results in nonresponse survey research (Johnson et al., 2002). Minority group oppression represents feelings of mistrust and resentment towards a majority group. Barriers to the advancement of these groups "such as poor education, poverty, and discrimination" may also translate to barriers in the survey process (p. 66). Helping behaviors exemplify more community-based attitudes; in survey research, people with these behaviors are more likely to participate or respond. Urban locations are less likely to exhibit these helping

behaviors (Groves & Couper, 1998; Porter & Umbach, 2006a). In this study, indicators of culture demonstrated patterns in SET response.

Hypothesis Five

Survey salience will increase the likelihood of response. Past research has demonstrated that survey participation is more likely if the individual sees the survey design and benefits as appealing (Groves et al., 2004; Groves et al., 2006; Groves & Peytcheva, 2008; Groves, Presser, et al., 2004; Heberlein & Baumgartner, 1978; Remedios & Lieberman, 2008). As implied, what interests one person may not interest another, and potential respondents weigh their preferences and influences in their decision to respond or decline.

There are several ways in which the variables of this study measured salience. First, the student's choice of major represented an area of interest. If the course taken was in the same department as the academic major of the student, then according to the theory of salience, the likelihood of response would have increased. However, students in majors outside of that department may not complete the evaluation because the topic did not apply to or interest them. Second, a student's grade in the course may also make the SET more salient. If they could associate that evaluation with something good, in this case a high grade, SET submission may have been more likely (Avery et al., 2006; Clarksberg et al., 2008; Dey, 1997; Fidelman, 2007; Porter & Umbach, 2006a; Sax et al., 2003, 2008).

Methods of Analysis

Using these hypotheses linked to nonresponse theory and the research questions for this study, I analyzed the data in efforts to predict the likelihood of response to SETs. This discussion consists of three major parts. First, an overview of the technique explains why this modeling procedure was useful to the purpose of this study. Second, I identify predictors of the dependent measure according to their specific hierarchical level. Third, the expected procedures section describes aspects of the modeling process that were integral to explaining the most variance in nonresponse.

Techniques

I conducted the analyses in two stages. In the first stage, descriptive analyses explored differences between respondents and nonrespondents. This provided an overview of the participants involved in this study, set the foundation to understand better the data, and began to uncover potential areas of nonresponse bias.

The second set of analyses employed a mixed model technique called hierarchical linear modeling (HLM) or multilevel modeling (MLM). Typically, regression models predict or explain variance in the dependent variable, which in this case is binary – whether the student submitted the online SET. However, an essential assumption to regression techniques is the independence of observations; each observation should have no relationship to another. Educational research rarely sees independent observations, more commonly encountering data structures that are naturally hierarchical (Allison, 1999; Groves & Couper, 1998; Porter & Umbach, 2006a; Raudenbush & Bryk, 2002; West et al., 2007). For example, in higher education students can be nested in classes,

departments, programs, colleges/schools, and institutions. Institutions are nested within regions, conferences, or divisions. While techniques in standard ordinary least squares (OLS) regression analyses are available to transform or combine data to account for the lack of independent observations, these techniques require the dismissal of factors that potentially explain the dependent variable. Instead, MLM improves the estimation of individual effects (Raudenbush & Bryk, 2002).

Several studies on survey nonresponse have incorporated multilevel modeling techniques. Porter and Umbach (2006) focused on a national survey, where data nested students within institutions. Hox and de Leeuw (2002) discovered that when predicting response rates for dichotomous outcomes, logistic multilevel models exhibited a much better fit than standard logistic regression modeling. Groves and Couper (1998) employed multilevel modeling, which they also called random-effect models, when studying nonresponse to surveys administered by the U.S. Census Bureau. They chose this technique because "households within the same neighborhoods share similar local ecological influences", and observations failed to meet the assumption of independence (p.58).

Likewise, the data in this study did not consist of independent observations. Instead, clusters of individual evaluations are within students (Figure 3.1). Since each of the 134,929 cases represented an SET request, most students would have received more than one evaluation. For each student, every class that was eligible for a course evaluation submission is a level-one case. If a student took five classes, five cases link to the one student. Participants in multiple observations within the dataset demonstrate a

violation of standard regression assumptions. To be able to incorporate both the individual and group predictors within the dataset, MLM accounts for dependent observations by modeling the affects of the nested data's individual effects.

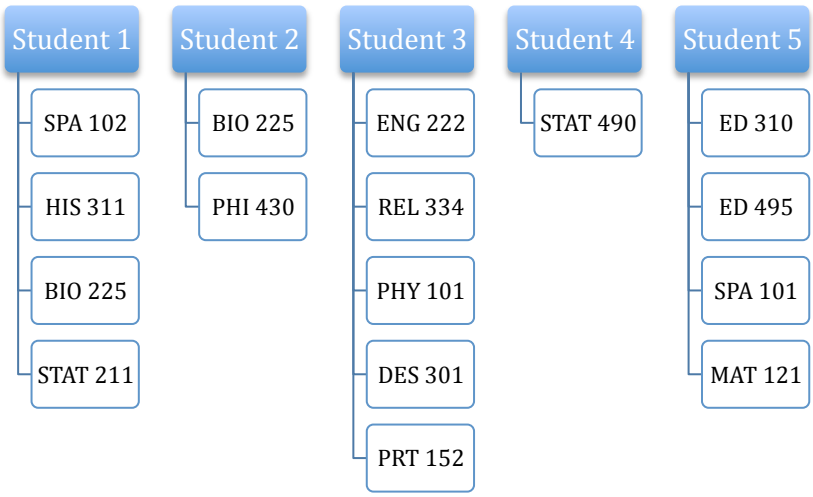


Figure 3.1. **Nested data structure in course evaluations.**

MLM employs the use of levels to separate the nested data (level-one) from the nest (level-two). That is, level-two contains the groups or clusters in which the observations reside. Level-one predictors describe and pertain to each observation. In this study, the 134,929 SETs administered to the students comprised level-one, and the 22,639 undergraduate students comprised level-two. The figure above represents these

two levels and the nested data structure. Table 3.3 lists variables available for this study and describes their measurements.

Table 3.3. **List and measurement of variables.**

Continuous Variables	*Non-Continuous Variables*
Number of Semesters Enrolled	International Student (dummy variable)
Number of SETs Administered	Entry Semester (several levels)
Total SCH	Gender (male or female)
SAT score (600-1600 range)	Race (several levels; dummy variables created)
NC State SCH	Semester GPA (4-point scale)
Semester SCH	Part-time status
	Transfer student (y/n)
	Personality/Environment Type (dummy variables created)
	Residency (y/n)
	Age (traditional or non-traditional)
	Took TOEFL (y/n)
	Student athlete (y/n)
	Campus housing (y/n)

Level-One Predictors.

Level-one consists of individual observations nested within level-two (Garson, 2009); that is, level-one variables are the contents of the "nest". Here, the observations at level-one were the individual SETs sent to the students. The dataset contained 134,929 SETs submitted to undergraduates for completion. Variables associated with each course evaluation are: 1) if the course was in the same department as the major and 2) the grade earned for that course.

Level-Two Predictors.

The variables at the second level are the groups that cause the observations to be interrelated (Garson, 1999); that is, level-two variables are the nests that contain or hold the contents. The nest in this case refers to the students, and level-two variables are characteristics attributed to them.

Analytical Procedures

MLM is based on a linear regression equation with a continuous dependent variable (Luke, 2004; Raudenbush & Bryk, 2002). However, this study used a binary dependent variable, also called a Bernoulli distribution. Because the outcome was dichotomous with 0 = no submission or nonresponse and 1 = response, a logistic multilevel model was necessary. HLM software, specifically developed for multilevel modeling, provides a logistic or binary equation (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004; Tabachnick & Fidell, 2007). For this study I used HLM 6, developed by Raudenbush, Bryk, and Congdon, to conduct analyses. However, HLM 6 does not allow for ordinal variable use or recoding (West et al., 2007; Garson, 2009). Dummy codes and

other related variables were created in SPSS for transfer into the HLM program. The screenshot below (Figure 3.2) displays a multilevel model in development using the HLM 6 program.

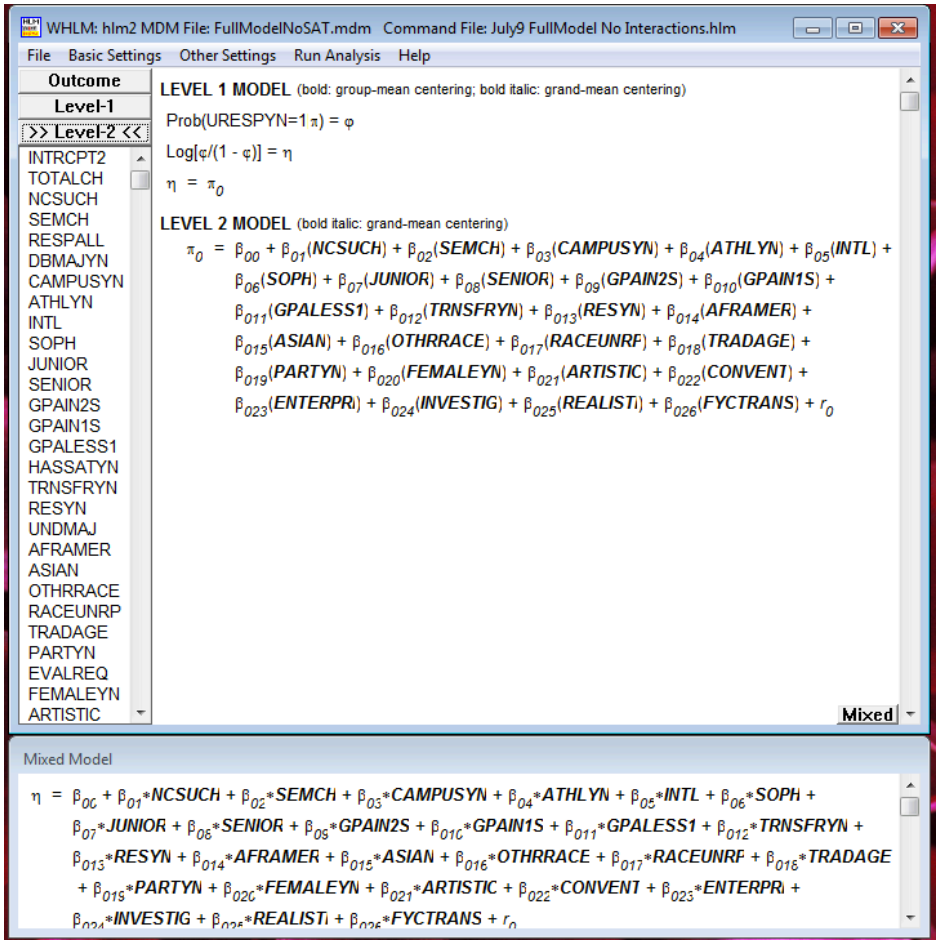


Figure 3.2. **HLM software screenshot.**

Standard logistic regression techniques fix the slopes and intercepts, meaning that the regression equation never accounts for characteristics associated with level-two predictors seen in MLM. If using standard logistic regression in this study, I would have had to ignore one of the two levels of variables that could have predicted the likelihood of response to avoid multicollinearity and large standard error. However, theories of nonresponse indicate that variables at both levels are possible explanations of nonresponse. To address these issues, a random-effects multilevel model employs additional sources of variability from both levels of predictors, interpreting the means as outcomes (Heck & Thomas, 2000; Luke, 2004). A fully random model incorporated into the equation both the means and the slopes as outcomes and helped to determine interactions across the two levels (Tabachnick & Fidell, 2007). This cross-level interaction factors in level-one and level-two predictors, and in this case I tested interactions of course and student level variables on course evaluation nonresponse.

Centering is much more common in MLM than other modeling techniques (Heck & Thomas, 2000; Luke, 2004). The process of centering subtracts the mean from the value of each predictor, changing the value to a deviation. The interpretation of the intercept thus changes; it is no longer the value of zero and is instead the mean of the dependent variable(s). Centering is useful to account for outliers and highly correlated predictors (Luke, 2004; Tabachnick & Fidell, 2007). Centering level-one variables is almost always recommended, and level-two centering is also recommended as it keeps consistent the interpretation of the results. In this case, I grand-mean centered all variables, to maintain consistency within the models.

The results in chapter four are from output of population-average models with robust standard errors, unless otherwise specified. With logistic multilevel analyses, four types of output usually appear: unit-specific, unit-specific with robust standard errors, population-average, and population-average with robust standard errors. Models with robust standard errors tend to increase the generalization of the results by reducing the likelihood of type I error and accounting for outliers and uneven data structures (as displayed in Figure 3.1 with the different number of SETs each student may receive to complete). Population-average models are only displayed in non-linear HLM output, which is the case here. Unit-specific models are parallel to the output displayed in linear models (Raudenbush et al., 2002), and do not address the difference across all possible values as population-average models do. Therefore, population-average models are appropriate.

HLM provides two major statistics to interpret model fit and reliability, because "traditional methods for variance-covariance component estimation fail to yield efficient estimates" (Raudenbush et al., 2002, p. 10). The tau statistic reports the variance between the students (level-two). The intercept reliability measures how reliable the model is on a scale of 0.000-1.000. When necessary, these two statistics compared models to determine the best fit.

Limitations

Several limitations of this study are evident. First, the setting excluded other institutions. Including other colleges and universities would have provided insight into other student populations and increased the generalizeability of the results. However, this design increased the relevance of the study to the institution.

Second, the exclusion of graduate students prevented their response probability from analysis. However, the inclusion of graduate students may have decreased the generalizeability of findings because not all postsecondary institutions have graduate programs. Also, this population may be characteristically different from undergraduates, and the data available for them would not be consistent if measured against the bachelor's degree-seeking students.

Third, the dataset excluded all faculty and most course characteristics. Since personnel decisions consider results from SETs, the results are also part of individual faculty member's employee records. Therefore the dataset excluded any characteristics potentially leading to the identification of instructors. Having this information would have increased the availability of level-one predictors in the analysis.

Missing data was not a large issue with this dataset. Since access to the course evaluation system requires the university-wide username and password, variables were mostly complete. However, SAT score was the one variable missing from several students' records. The next chapter further addresses the handling and modifications for missing data.

Summary

Because of the hierarchical structure of data, a multilevel model was the best fit for analysis. This technique accounts for the lack of independence of observations and the clustering of data at two levels. Additionally, in this study MLM also supported theories of nonresponse, which presented survey cooperation as a multi-faceted decision with several considerations influencing participation. These analyses described the likelihood of SET submission at NC State.

A study of this magnitude on course evaluation nonresponse is the first of its kind. Most studies have only considered nonresponse in a handful of classes, took place in private settings, and did not use as many variables as are available to this study (Fidelman, 2007; Ho & Shapiro, 2008; Jones, 2009; McGourty et al., 2002a, 2002b; Thorpe, 2002). Moreover, much is at stake when considering SETs in this context; administrators make personnel and academic decisions with the results in consideration. In order to improve data quality and make better-informed administrative decisions, this study sought to discover factors of SET participation and the likelihood of response.

CHAPTER FOUR: RESULTS

Postsecondary institutions are increasingly changing the method of administration and data collection for course evaluations (student evaluations of teaching, or SETs) from paper surveys to an online format. With this implementation comes the concern of lowered response rates and maintaining high quality data obtained from the SET results. To investigate nonresponse to online SETs, this study explores variables that describe the student as well as the course for which the evaluation was administered.

This chapter describes the results from data analyses. First, an overview of the data and preliminary analyses provides a foundation to understanding the data. Second, multilevel models demonstrate results according to the five hypotheses in chapter three. Lastly, an overarching model delivers results across variables.

Descriptive Analyses

Level-One Variables

The University Planning and Analysis (UPA) office at North Carolina State University (NC State) administered 134,929 SETs using the ClassEval instrument in the fall semester of 2009, 48.9% of which were partially or fully completed and submitted. On average, students received about six SETs each, with a range of 1-23. Table 4.1 describes key information about courses

Table 4.1. **Level-one descriptive statistics: SETs.**

Variable	*N*	*Mean*	*SD*
SETs Administered	134,929	5.96	2.02
SETs Submitted*	66,024	0.49	-----
Course in Dept. of Student's Major	134,929	0.27	0.45
Grade in Course of SET			
Grade = A	44,832	0.33	0.47
Grade = B	39,001	0.29	0.45
Grade = C	17,796	0.13	0.34
Grade = D	4,832	0.04	0.19
Grade = F/U	5,926	0.04	0.20
Grade = S	7,123	0.05	0.22
No Grade**	15,419	0.11	0.32
Average Course Grade* **			
A	846	0.22	0.41
B	1,763	0.45	0.50
C or Below	707	0.18	0.39
Course Not Graded	578	0.15	0.36

* See Table 4.3 for characteristics of submitted SETs
** Course was ungraded or the student received an I, AU, W, or other non-grade.
*** Course did not award GPA points. Includes: Pass/Fail and non-graded courses. Average grade for class (3894 courses).

Level-Two Variables

This study included all undergraduate students and their undergraduate courses. Table 4.2 provides an overview of the 22,639 students at NC State in these categories, and Appendix J provides additional descriptive statistics for those level-two variables not listed below. The student population on average received about six evaluations for the Fall 2009 semester, was 56.2% male, and averaged 14.53 credit hours per student. In general, students appear evenly distributed by class. Only 15.9% of undergraduates transferred from another institution and about 33% live on campus.

Table 4.2. **Level-two descriptive statistics: Students.**

Variable	*N*	*Mean*	*SD*
Students	22,639	-----	-----
Female	9927	0.44	0.50
SETs (per Student)	134,929 (1-23*)	5.96	2.02
Credits			
Semester Credit Hours (CH)	0-23*	14.53	2.61
Total CH	0-355*	64.36	43.64
NCSU CH	0-301*	62.28	42.01
Part-Time Students	1,264	0.056	0.23
Race *(White/Caucasian = ref.)*			
African American	1,957	0.09	0.28
Asian	1,230	0.05	0.23
Did Not Disclose	1,070	0.05	0.21
Other	859	0.04	0.19
Class Rank *(First Year = ref.)*			
Sophomore	5,177	0.23	0.42
Junior	5,383	0.24	0.43
Senior	6,420	0.28	0.45
Semester GPA *(3.0-4.0 = ref.)*			
GPA 2.0-2.99	9,426	0.42	0.49
GPA 1.0-1.99	1,363	0.06	0.24
GPA 0.0-0.99	208	0.09	0.10
Major Environment *(Social = ref.)*			
Artistic	620	0.03	0.16
Conventional	433	0.02	0.14
Enterprising	3,536	0.16	0.36
Investigative	10,512	0.46	0.50
Realistic	4,256	0.19	0.39
FYC/Transitions	1,413	0.06	0.24
SAT score	600-1600*	1179.96	134.35
Transfer Student	19,302	0.16	0.37
On Campus Housing	7,429	0.328	0.47
International Student	226	0.01	0.10
NC Resident	20,736	0.92	0.28
Part-Time	1,264	0.06	0.23
Traditional Age (<25 years)	21,208	0.94	0.24
Athlete	596	0.03	0.16

* Indicates Range, not N

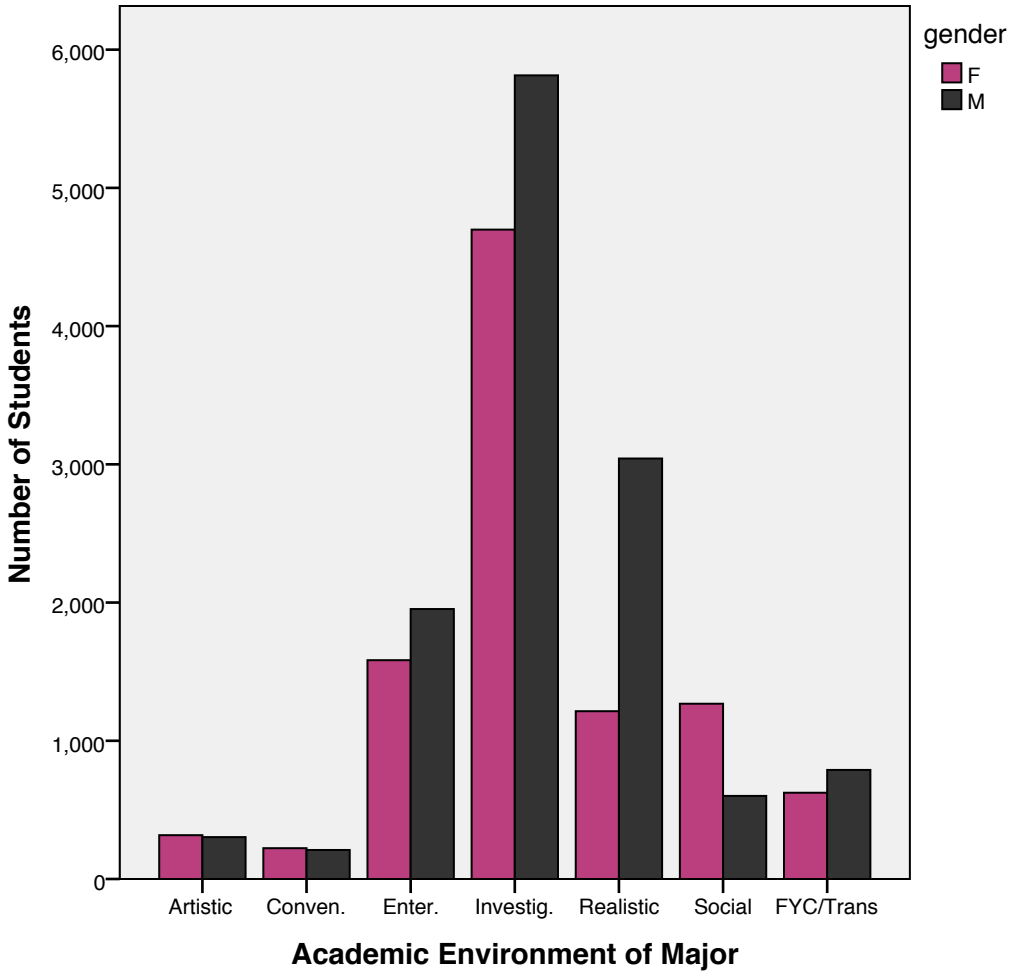


Figure 4.1. **Gender of students in each academic environment.**

Appendix I links each of the six academic discipline types to the majors at NC State. Figure 4.1 displays the gender of students in each academic environment, as measured by major (Holland, 1974; Smart et al., 2000). Because there were almost 3,000 more males than females at NC State, it is not surprising to see that most categories have

more males. Realistic environments appear to have had the largest male to female ratio. However, social environments contained more females than males, seen also in artistic and conventional environments, though the differences appear small and therefore affecting few students.

Primary Analyses of SET Participation

Table 4.3 describes the response rates for the 22,639 students and 134,929 ClassEvals administered. It displays how many completed all, none, or only some of their SETs, and what percentage of our student population fits into each category. The rate of SET completion represents the number of SETs submitted by students divided by the number SETs administered. Most students either responded to all (n=10,958) or responded to none (n=10,313). Over 5% of undergraduates (n=1,368) were partial respondents, submitting some of SETs, but not all.

Table 4.3. **SET response rates: All or nothing?**

Rate of SET Completion	*Frequency*	*Percentage of N*
0 (Completed No SETs)	10313	45.6
Completed some, not all		
0.01-0.09	13	0
0.10-0.19	257	1.1
0.20-0.29	250	1.1
0.30-0.39	121	0.5
0.40-0.49	125	0.6
0.50-0.59	192	0.8
0.60-0.69	136	0.6
0.70-0.79	92	0.3
0.80-0.89	187	0.8
0.90-0.99	8	0
1 (Completed All SETs)	10958	48.4
N (number of students)	22639	

In Figure 4.2, the grade earned in the course reflects the average response rate for each grade category. Grades emerged as a potential predictor of response. Though students complete their evaluations before the final exam period begins, thus before getting their final grades, there is still a noticeable correlation between those that received higher grades and others. Response rates for As were over 58%, while Ds, Fs, and No Grades were lower than 40%.

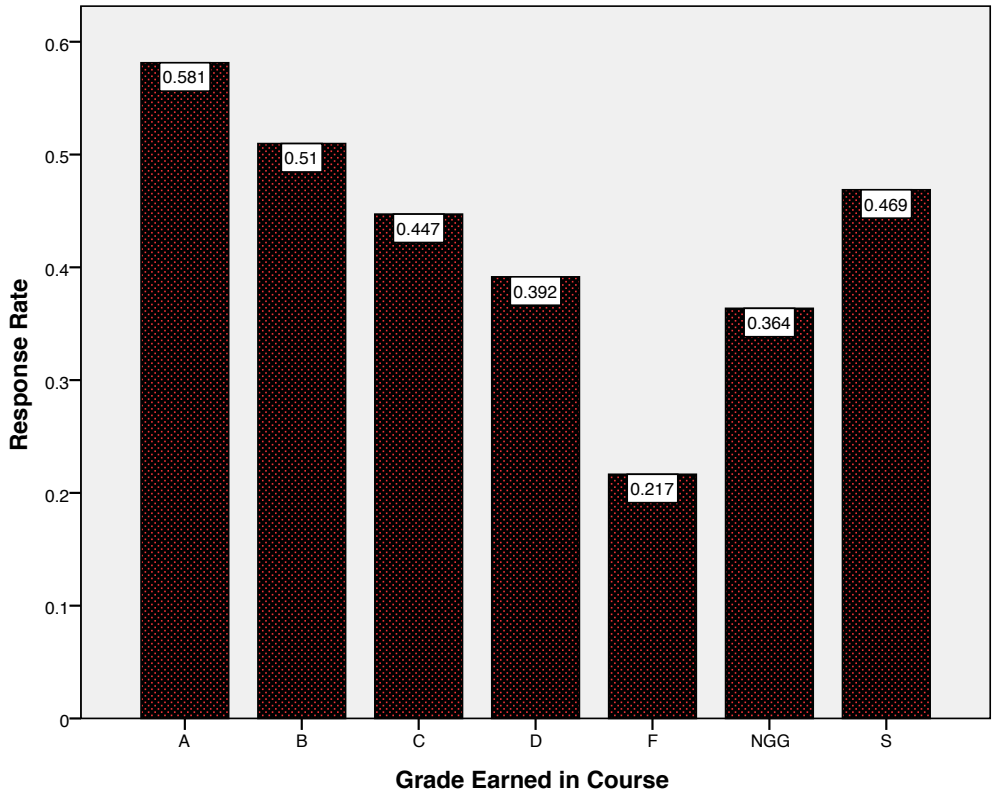


Figure 4.2. **SET submission rate by grade earned.**

Additional characteristics associated with submitted SETs are in Table 4.4. Both level-one and level-two variables comprise this table, but it is important to note that this cross-tabs analysis did not take into account the nested data structure and lack of independence. These statistics represent SET submission characteristics at both levels: the student who received the SET and the course of the SET.

Table 4.4. **Characteristics of submitted SETs.**

Variable	*Frequency*	*Response Rate*
Commuters	39,700	0.46
On Campus Housing	26,324	0.54
Non-Athletes	64,729	0.49
Athletes	1,295	0.36
Race		
African American	4,937	0.43
Asian	3,464	0.47
Race Not Reported	3,334	0.52
Other	2,372	0.46
White	51,917	0.50
Female	32,320	0.54
Male	33,704	0.45
NC Resident	60,025	0.49
Not NC Resident	5,999	0.50
Major Environment		
Artistic	1,639	0.52
Conventional	1,157	0.51
Enterprising	8,953	0.46
Investigative	31,163	0.48
Realistic	13,213	0.51
Social	5,300	0.49
No Environment*	4,599	0.51
Full-Time Student	64,640	0.49
Part-time Student	1,384	0.41
Not Traditional Age	3,210	0.52
Traditional Age	62,814	0.49
Not Transfer	56,396	0.49
Transfer Student	9,628	0.49
Class Rank		
First Year	20,594	0.54
Junior	14,364	0.46
Sophomore	14,330	0.46
Senior	16,736	0.48
Grade Earned		
A	26,065	0.58
B	19,880	0.51
C	7,958	0.45
D	1,892	0.39
F	1,283	0.22
S	3,339	0.47
No Grade	5,607	0.36
Course Not in Major Dept	49,044	0.49
Course In Major Dept	16,980	0.49

* First Year College (FYC) and Transitions Students are Undecided Majors.

Differences in response rates emerged for several variables when compared to the university's response rate for Fall 2009 (48.9%). Response rates were below 40% for SETs administered to grades of D, no grades, and athletes. SETs associated with grades of F were the lowest at a 22% response rate. The characteristics with the highest response rate, all of which were higher than university's average response rate, were the following: grades of A (58%), first-year students, students living on-campus, and females (all with 54% response rates).

Of particular interest is the course location variable. There was almost no difference in SET response rates for courses in or outside of the department of the student's major (Table 4.4). However, with the separation of course locations by the academic environment of the student's major, differences across environment types become evident (Figure 4.3).

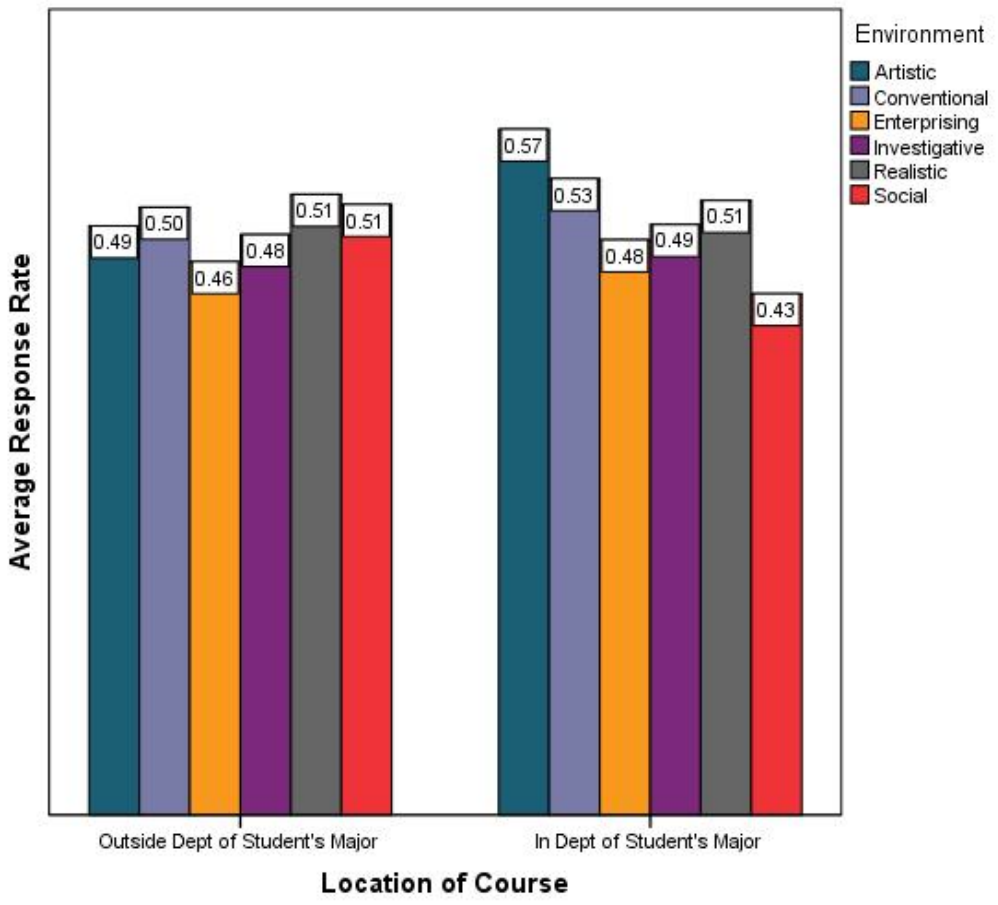


Figure 4.3. **Academic and course environment: Response rates.**

Remembering that the response rate was higher for SETs submitted by females versus males (Table 4.4) and that there were more females in social, artistic, and conventional disciplines (Figure 4.1), it might be natural to infer that these environments would have had higher response rates. In addition, there were no apparent differences in response rate

if the course was in the same department as the student's major (Table 4.4). Nevertheless, the separation of the student's personal and major environment and their course's location (Figure 4.3) revealed some differences in response rates. Artistic and social types changed the most if the course was in the same department as their major. The response rate decreased for social types if the course and major matched departments, while the response rate for artistic types increased with the same interaction.

Academic environment and grades graphed in Figure 4.6 display these groups' differences in response rates. Overall, grades were consistent predictors of response, despite the academic environment of the student's major. Figure 4.6 also displays the overall consistency of grades as a predictor of SET response. Some variance with conventional, artistic, and no environments (undecided majors) was evident from lines crossing or inconsistent paths based on other groups. Grades also varied some when accounting for environment of the student's major – S, C, D, and No Grades in particular. However, these graphs did not account for the nested data structure within variables. Further analyses in multilevel models accounted for the nested SETs within students.

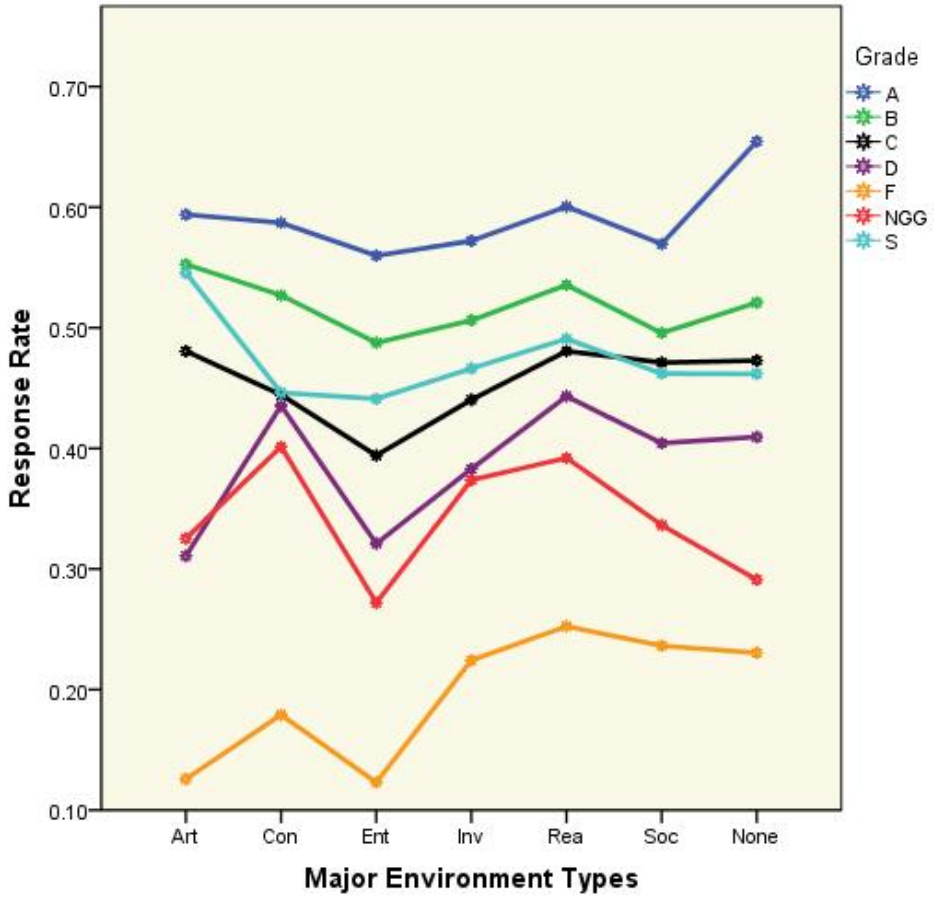


Figure 4.4. **Response by major environment: Does grade make a difference?**

Finally, when considering survey fatigue and the number of SETs a student needs to complete, what is the response rate? Did the number of SETs administered to a student influence their likelihood of responding?

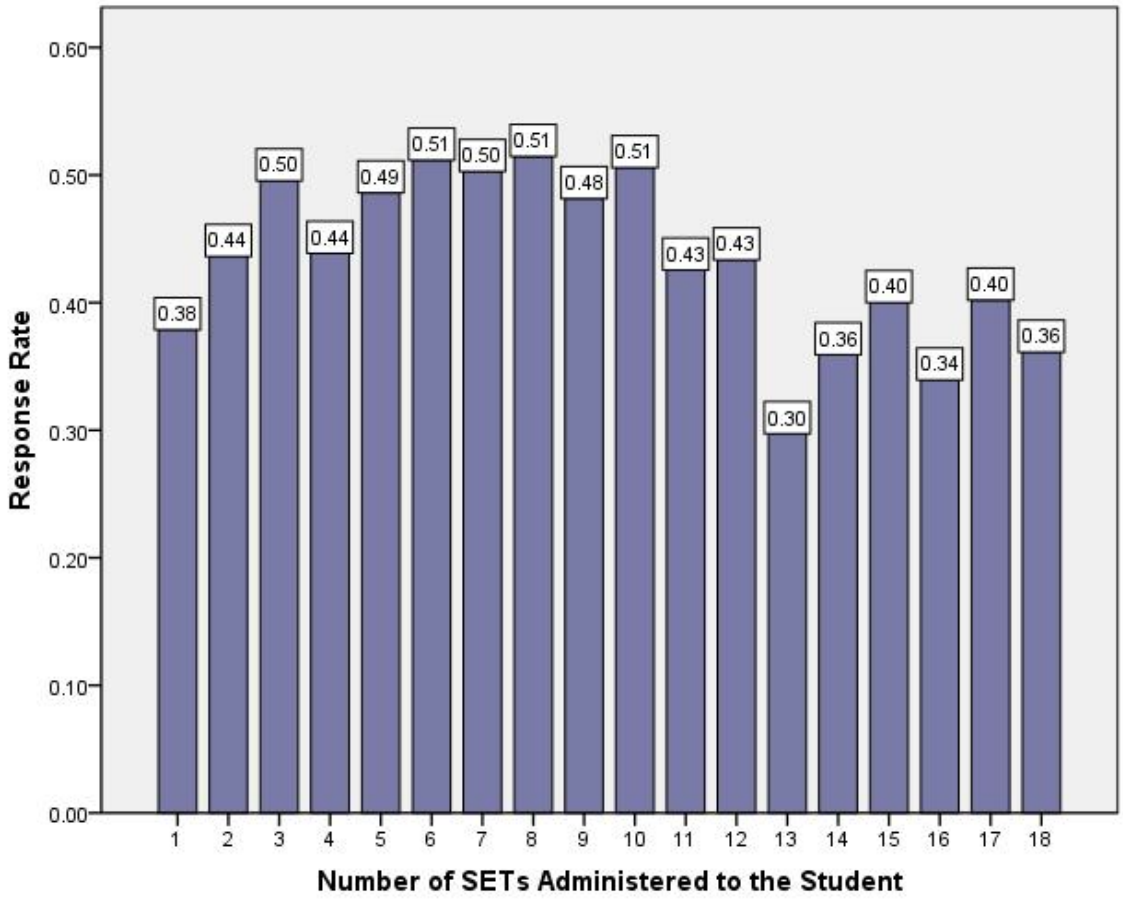


Figure 4.5. **Too many SETs? Response rate by SETs administered.**

In Figure 4.5 a noticeable decrease in response rates appeared if there were 11 or more SETs administered to the student. Similar decreases appear in graphs located in the Appendix J. This figure may indicate that survey fatigue began when a student received 11 or more SETs to complete in one semester, because submission rates were over five

percentage points lower than the average SET response rate. One, two, or four SETs also demonstrated lower response rates. Further analyses considered possible reasons for low response rates on both ends. Multilevel models explored reasons for these differences in opportunity costs, survey fatigue, and culture hypotheses and theories.

Multilevel Models: An Overview

With a Bernoulli distribution, that is, a binary outcome of 0 or 1, logistic multilevel models were employed, and an appropriate software to use for this type of analysis is HLM 6 (Luke, 2004; Raudenbush et al., 2004; Tabachnick & Fidell, 2007). Multilevel modeling also better displays interactions between level-one and level-two variables. Where no differences were evident when performing analyses in ordinary least squares regression (OLS) or correlation procedures, hierarchical modeling was able to explore further predictor interactions at both levels. For example, as displayed above in Figure 4.4, there appears to be no difference in response rates for course location. However, within course location academic environment displayed some differences in Figure 4.5. Similar to these graphs, the tables below display, when appropriate, interactions between level-one and level-two variables for each model. I developed models according to the variables in each hypothesis, randomizing intercepts and slopes and using both as outcomes in the models (Luke, 2004).

The tables below each display at least two models, the first model includes the variables related to the specific hypothesis and the other presents coefficients from the fully controlled model (displayed in Appendix K). For example, hypothesis one addressed demographics only, but the controlled model included demographics as well as

housing, residency, class rank, and all other variables. Tables for hypotheses four and five contain three models. Because interactions were appropriate, the table displays a main effect model (for the specific hypothesis), a fully controlled model without interactions, and a fully controlled model with interactions.

Hypotheses Testing and Results

Hypothesis One

Females, white students, and older students will be more likely to respond to SETs than their peers. NC State is unlike many other universities because of the larger male population (n = 12,712, 56.2%). Of the undergraduate student body, 77.4% report to be of a Caucasian background, 8.6% were African-American, 5.4% reported their race as Asian, and the remainder is unknown. Students at NC State were mostly 25 years old or younger. Only 6.3% of the undergraduate student population was 26 years and older. Table 4.2 contains additional information on demographics.

To determine the effects of these student-level variables on the likelihood of SET response, a multilevel analysis was performed. Variables were grand-mean centered, meaning that coefficients are calculated based on the mean. In this case, only a unit-specific model with robust standard errors is available because hypothesis one contained no level-two predictors. None of these variables were continuous; the creation of dummy variables allowed the results to be interpreted against a reference group. In Table 4.4, males, students of white/Caucasian race, and non-traditional age students (defined as 26 years and older) are the reference groups.

Table 4.5. **Demographics: Uncontrolled and fully controlled models.**

Variable	*Demographics Only*		*Controlled*	
	Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Intercept	-0.048*	0.953	-0.063*	0.939
Female	0.655***	1.925	0.351***	1.421
Race				
African-American	-0.546***	0.579	-0.226***	0.798
Asian	-0.192*	0.825	-0.154**	0.857
Other Race	-0.216*	0.806	-0.111	0.895
Race Unknown	0.139	1.148	0.007	1.007
Traditional Age	-0.301***	0.740	-0.362***	0.696
Tau	6.302		6.059	
Intercept Reliability	0.750		0.735	

*** p<0.001, **p<0.01, *p<0.05

The coefficient represents the change for the variable against the reference groups, and because variables were grand-mean centered, the change is measured based on the group mean. The coefficients are also interpreted based on the distribution of the outcome, in this case (with a logistic multilevel analysis) a binary distribution with “1” representing response and “0” representing nonresponse. The odds ratio describes the likelihood of the outcome. An odds ratio (OR) statistic of two can be interpreted that the outcome is twice as likely, while a score of 0.50 describes the outcome as half as likely. In the fully controlled model (left side), females were 1.42 times as likely to submit an SET. Those who reported their race as African-American, Asian, and Other were significantly less likely to respond than Caucasian/White students. There was no significant difference between the reference group and those who did not report their

race. Finally, students 25 years of age or younger were less likely to respond than non-traditional age (OR = 0.70).

The full model included the demographics as well as the other variables. When comparing the two, the statistics changed, but most remained as significant as the hypothesis one model. For example, the coefficients for female and African Americans dropped almost half when incorporated into the full model, but remained just as significant. Variables for traditional age and Asian categories changed very slightly, but again remained just as significant. The only change in significance was the other race category; it was not significant in the full model.

Hypothesis Two

High academic performance will increase the likelihood of SET response. As previously described, academic performance measures include both level-one and level-two variables. The grade earned in the course and the average grade for the class were two level-one variables, and the level two variables were SAT score and dummy variables based on the student's GPA for the semester.

Grades of "A" were awarded in over 33% of the 134,929 cases. "B" grades comprised of 28.9% of the cases, "C" grades 13.2%, and 11.4% of the 134,929 observations received no grade, an incomplete, or a withdrawal. "S" grades (5.3%) are satisfactory grades, typically found in a pass/fail class. Their counterparts, unsatisfactory grades, combined with failing grades since, in both cases, the student did not complete the requirements for credit (4.4%). Table 4.1 has further information on level-one descriptive statistics.

At level-two (Table 4.2), two variables measured academic performance: the student's semester GPA for Fall 2009 based on a 4.0 scale and SAT score ranging from 600-1600, measured in 100-point increments. SAT score was not available for all students for the following three reasons: transfer students do not typically submit SAT scores, some students submit alternatives to SAT scores, and some students petition to waive this requirement for acceptance. To explore the missing data, I examined several different models for potential differences. What would happen if students with SAT scores were omitted from the dataset? How did they compare to the models with SAT scores? Were there any differences in models excluding and including transfers? Tables 4.6 and 4.7 report results from models with and without SAT scores, and displayed no significant differences. There are none or minor differences seen in Tables 4.6 and 4.7, as well as in Appendix M when comparing models with and without the SAT variable.

When considering methods for imputation, the mean score of 1180 would be problematic when considering transfer students, students older than 25 years, and other reasons for the missing data. Transfer students and the presence of the SAT were highly, significantly, and negatively correlated. However, transfer status is not appropriate in an academic performance model. Instead, hypothesis four addressed transfer students in the form of culture. Alternatively, none of the variables appropriate for imputing SAT score, which would be similar achievement variables, were available to this study (e.g. overall GPA, ACT score, high school GPA). List wise deletion is widely practiced for missing data if comparison models are available and if other variables (e.g. transfer status) are available to account for the deleted cases (Allison, 1999). Therefore, I present two

models below – one including all students but excluding the SAT variable, and a second model excluding students without an SAT score and including the SAT variable. I used the transfer student variable in all full models and in hypothesis four.

To determine if academic achievement influenced the decision to participate in the SET submission process, four logistic multilevel models examined achievement. Semester GPA was a constant variable throughout the four models. These models, displayed in two tables, include all students and did not use SAT score as a predictor (n=134,929 SET submissions) in addition to models that excluded students who were missing the SAT score (n=119,190 SET submissions). Appendix L explores interactions between academic achievement variable levels, but overall displays the same results as models without interactions. Appendix M explored interactions of the average class grade and also contains SAT and non-SAT models.

Table 4.6. **Academic achievement of all students: Uncontrolled and fully controlled models.**

Variable		***Academic Achievement***		***Controlled***	
		Coeff.	*Odds ratio*	*Coeff.*	*Odds ratio*
Level One *(Grade A = ref.)*	Intercept	-0.056***	0.945	-0.063*	0.939
Student Grade = B		-0.029*	0.971	-0.004	0.996
Student Grade = C		-0.093***	0.912	-0.038**	0.963
Student Grade = D		-0.149***	0.861	-0.064**	0.938
Student Grade = F/U		-0.350***	0.705	-0.422***	0.656
Student Grade = Satisfactory (S)		-0.531***	0.588	-0.351***	0.704
No Grade		-0.505***	0.603	-0.488***	0.614
Level Two *(GPA 3.0-4.0 = ref.)*	Intercept	-0.072***	0.930	-0.063*	0.939
Semester GPA in 2s		-0.498***	0.608	-0.446***	0.640
Semester GPA in 1s		-1.135***	0.321	-1.198***	0.302
Semester GPA less than 1		-1.903***	0.149	-2.230***	0.108
	Tau	6.354		6.059	
	Intercept Reliability	0.743		0.735	

*** p<0.001, **p<0.01, *p<0.05

"No Grade" and grades of F and S were less likely to respond when compared to that of an A grade earned. This is not dissimilar to Figure 4.2. At level-two, a lower semester GPA also influenced the likelihood of response. When compared to a 3.0-4.0 GPA, students with lower grades for that semester were significantly less likely to respond. Students with a semester GPA between 2.0 and 2.9 were 0.61 times as likely as the reference group to complete an SET, with 1.0-1.9 GPAs (OR = 0.32) and GPAs less than 1.0 (OR = 0.15) continued the pattern of the probability of nonresponse. Overall, the models in Tables 4.6 and 4.7 produced no significant changes between them, though grades of B, C, and D display slight changes in ρ-values.

Table 4.7. **Academic achievement: Uncontrolled and fully controlled models.**

Variable		***Academic Achievement***		***Controlled***	
		Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Level One *(Grade A = ref.)*	Intercept	-0.072***	0.930	-0.066***	0.937
Student Grade = B		-0.021*	0.980	0.003	1.003
Student Grade = C		-0.080***	0.923	-0.029*	0.972
Student Grade = D		-0.136***	0.873	-0.056**	0.946
Student Grade = F/U		-0.520***	0.594	-0.412***	0.662
Student Grade = Satisfactory (S)		-0.343***	0.709	-0.348***	0.706
No Grade		-0.489***	0.613	-0.476***	0.621
Level Two *(GPA 3.0-4.0 = ref.)*	Intercept	-0.089***	0.914	-0.091***	0.913
Semester GPA in 2s		-0.533***	0.587	-0.446***	0.640
Semester GPA in 1s		-1.224***	0.294	-1.156***	0.315
Semester GPA less than 1		-1.979***	0.138	-2.083***	0.125
SAT score		-0.001***	1.000	-0.001***	0.999
	Tau	6.397		6.274	
	Intercept Reliability	0.747		0.739	

*** p<0.001, **p<0.01, *p<0.05

As evident from these models, there were few, if any, differences between the models that included or did not include SAT scores (see also Appendices L and M). Whether the model included or excluded SAT score, coefficients remained the same or changed slightly and insignificantly. Where there was statistical significance in one model for one variable or its interactions, statistical significance was also in the other. Controlled models (with all variables) did not immensely differ from the uncontrolled (hypothesis variables only) models. Grades of B, C, and D demonstrated slight changes in significance when comparing the two.

Additional models that incorporated the average course grade variable at level-one, the same level-two variables, and their interactions are in Appendix M. This reports the average grade for all students in that particular course. Available measures were class

grades of A (21.7% of courses), B (45.3%), C and under (18.2%), or No Grade (14.8%). Here, no grade meant that the course did not award grades or was a pass/fail course since these offer no GPA points. Similar results for A, B, C, and No Grade emerged when comparing these two models.

Hypothesis Three

Opportunity costs and survey fatigue will decrease the probability of response. To determine how students prioritized their efforts and if survey fatigue was an issue on this campus, I examined several variables attributable to these effects. In Table 4.2, first-year students consisted of 25% of the population, sophomores 23%, juniors 24%, and seniors 28%. Positive correlations between several of these variables indicative of survey fatigue existed. For example, the total number of credit hours is directly parallel to their status as a first-year, sophomore, junior, or senior. Also, the number of semester credit hours positively correlates to the number of SETs administered to the student for the Fall 2009 semester. I examined several models to determine which variables provided a better fit and were more reliable, the results of which are in Tables 4.8 and 4.9.

In the survey fatigue model (Table 4.8), class rank and the number of semester credit hours were both statistically significant. All three class ranks were less likely to respond when compared to first-year students. Sophomores and juniors responded similarly between models, but coefficients and odds ratios slightly increased for seniors. When included in the full model, seniors did not respond differently from first-year students. Appendix N contains results from models incorporating squared terms for total

credit hours and semester credit hours for survey fatigue. Statistics were significant but did not vary between squared-models or controlled and uncontrolled models.

***Table 4.8*. Survey fatigue: Uncontrolled and fully controlled models.**

Variable	***Survey Fatigue***		***Controlled***	
	Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Intercept	-0.051***	0.951	-0.060***	0.941
Class Rank *(First Year = ref. group)*				
Sophomore	-0.293***	0.746	-0.255***	0.775
Junior	-0.286***	0.751	-0.186**	0.830
Senior	-0.187***	0.829	-0.012	0.988
SETs Administered	0.001	1.001	-0.030***	0.971
SETs Administered (Squared)	-0.001*	0.999	-0.002***	0.998
Tau	6.353		6.047	
Intercept Reliability	0.752		0.735	

*** p<0.001, **p<.01, *p<.05

***Table 4.9*. Opportunity costs: Uncontrolled and fully controlled models.**

Variable	***Opportunity Costs***		***Controlled***	
	Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Intercept	-0.055**	0.947	-0.063*	0.939
Class Rank *(First Year = ref. group)*				
Sophomore	-0.208***	0.812	-0.247***	0.782
Junior	-0.146***	0.753	-0.183**	0.832
Senior	0.005	0.949	-0.013	0.987
On Campus Housing	0.249***	1.525	0.231***	1.260
Athlete	-0.501***	0.455	-0.382***	0.682
Part-Time	0.079	1.219	-0.041	0.960
Semester CH	0.042***	1.074	0.016*	1.016
Semester CH (Squared)	0.001***	1.001	0.000*	1.000
Tau	6.321		6.059	
Intercept Reliability	0.750		0.735	

*** p<0.001, **p<.01, *p<0.05

In Table 4.9, the opportunity costs model incorporated class ranks like survey fatigue, which also were statistically significant, but substituted semester credit hours for the number of SETs administered to the student. This model also included other variables representative of this theory, and part-time students, athletes, and students living on campus were also statistically significant. Several of these variables produced remarkable odds ratios. Athletes were 0.61 times as likely to respond compared to non-athletes (0.69 in controlled model), and students living on campus were 1.31 times as likely to respond as commuters (1.28 in controlled model).

Hypothesis Four

The probability of response will increase for students with stronger cultural and communal ties to the campus. Culture consciously and subconsciously affects all of our decisions, including the decision to participate, or in this case submit an SET for a class. The variables in the models below represent several aspects of personal and campus culture. Additional descriptive statistics for this hypothesis are in Appendix O, not Tables 4.1 and 4.2 like previous models. Descriptive statistics for this hypothesis were different because these models eliminated First Year College and Transitions students from the analyses. This was necessary because these two departments are undecided majors and do not sponsor or list any courses; therefore, the level-one variable lacked interactions with course location variable.

To evaluate potential cultural influences on nonresponse, I assessed two separate models for this hypothesis: academic environment (Table 4.10) and all other culture

groups (Table 4.11). This separation is for the following two reasons: first, this increases the reliability of the models, and second, I received error messages for high variable correlations when conducting interaction analyses between the two levels.

The social type was the reference group in the first model called Academic Environment (Table 4.10). When exploring the average likelihood of response in this uncontrolled model, there was no statistically significant difference. Only one exception was for enterprising types as they were 0.87 times as likely to respond compared to social majors.

Table 4.10. **Culture: Academic environment and full models.**

Variable	*Academic Environment*		*Controlled No Interactions*		*Controlled*	
	Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Level One						
Course Location (Int.)	0.125***	1.132	0.065***	1.067	0.125***	1.061
Level Two ✢ (Int.)	-0.056***	0.945	-0.067***	0.936	-0.069***	0.933
Artistic	0.042	1.043	0.047	1.048	0.054	1.056
Conventional	0.041	1.043	0.078	1.081	0.093	1.097
Enterprising	-0.145**	0.865	-0.045	0.855	-0.027	0.973
Investigative	-0.076	0.927	-0.004	0.996	0.015	1.015
Realistic	0.030	1.030	0.172***	1.188	0.188***	1.207
Interactions ✢✢ (Int.)	0.052***	1.053	-----	-----	0.056***	1.058
Artistic	0.240***	1.272	-----	-----	0.239***	1.269
Conventional	0.217**	1.242	-----	-----	0.215**	1.239
Enterprising	0.161***	1.175	-----	-----	0.155**	1.168
Investigative	0.143**	1.154	-----	-----	0.138**	1.148
Realistic	0.175***	1.191	-----	-----	0.161**	1.175
Tau	6.468		6.174		6.182	
Intercept Reliability	0.753		0.736		0.736	

*** p<0.001, **p<.01, *p<0.05
✢Social = Ref., ✢✢ Interacted with Course Location

In the Academic Environment model, when the major types interacted with the course location variable, they became more likely to respond at a statistically significant level. Even enterprising types, less likely to respond without interaction, were more likely to respond if the course was in the same department as their major. The figure below graphs the probability of response for academic environment of the student's major and of the course.

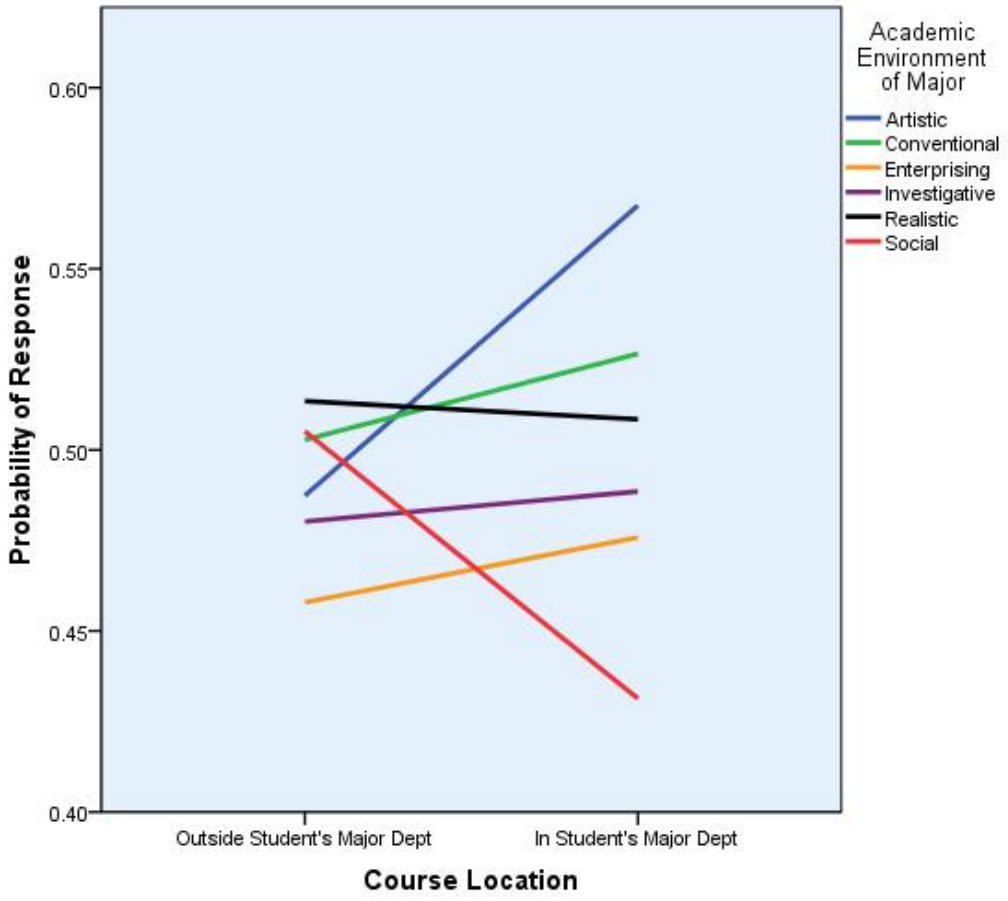


***Figure 4.6*. Probability of SET submission: Academic environments.**

In Table 4.1, response rates were not different for courses in or outside of the department of the student's major. However, disparities emerged when comparing the course environment to the student's major environment (Figure 4.6). Realistic majors

were more likely to submit an SET if the course location was outside the student's major, and enterprising majors were less likely to respond. If the department that housed the course also housed the student's major, artistic majors were more likely to respond, but social majors were less likely. Overall, SET submission was more probable for artistic, investigative, enterprising, and conventional majors if the course and student's major were in the same department. However, this trend was not consistent for social and realistic majors. While realistic majors did not change dramatically, there was a sharp decrease in the probability of response for social types. Most major types or environments were more likely to respond to an SET if the student's course and major were within the same department.

The fully controlled model, which contained all variables (Appendix O), contained similar outcomes especially with interactions. The enterprising category was not significantly different from the mean in the full model, and the realistic category, unlike the uncontrolled model, was significantly more likely to submit an SET.

Table 4.11. **Culture: Uncontrolled and fully controlled models.**

Variable		***All Culture***		***Controlled No Interactions***		***Controlled***	
		Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Level One							
Course Location	(Int.)	0.125***	1.132	0.065***	1.067	0.059***	1.061
Level Two	(Int.)	-0.058***	0.944	-0.067***	0.936	-0.069***	0.933
NCSU CH		-0.000	0.010	-0.002	0.998	-0.002	0.998
On Campus Housing		0.308***	1.361	0.214***	1.238	0.214***	1.239
Athlete		-0.330***	0.719	-0.273**	0.761	-0.277***	0.758
International Student		0.210	1.233	0.231	1.260	0.235	1.265
Transfer Student		0.089*	1.093	0.174***	1.190	0.178***	1.195
NC Resident		-0.041	0.960	0.024	1.024	0.024	1.025
Traditional Age (< 25 years)		-0.355***	0.701	-0.357***	0.700	-0.356***	0.701
Part-time		-0.339***	0.713	-0.040	0.961	-0.041	0.960
African American ✣		-0.306***	0.736	-0.211***	0.810	-0.211***	0.810
Asian ✣		-0.140**	0.869	-0.156**	0.855	-0.157**	0.855
Other Race ✣		-0.149*	0.853	-0.111	0.895	-0.112	0.894
Race Unknown ✣		0.011	1.011	0.002	1.002	0.001	1.001
Interaction (*CrsLoc) (Int)		0.065***	1.067	----	-----	0.057***	1.058
NCSU CH		0.000	1.000	----	-----	0.000	1.000
On Campus housing		-0.003	0.997	----	-----	0.008	1.008
Athlete		0.116	1.123	----	-----	0.131*	1.140
International Student		0.068	1.071	----	-----	0.061	1.063
Transfer Student		-0.046	0.955	----	-----	-0.066*	0.936
NC Resident		0.022	1.022	----	-----	0.020	1.021
Traditional Age (< 25 years)		-0.014	0.986	----	-----	-0.030	0.971
Part-time		0.016	1.016	----	-----	-0.015	0.985
African American ✣		0.013	1.013	----	-----	0.021	1.021
Asian ✣		-0.008	0.992	----	-----	-0.011	0.989
Other Race ✣		-0.001	0.999	----	-----	0.008	1.008
Race Unknown ✣		0.011	1.011	----	-----	0.011	1.011
Tau		6.426		6.174		6.182	
Intercept Reliability		0.750		0.736		0.736	

***p<0.001, ** p<0.01, *p<0.05

✣ Caucasian/White = reference

The second model incorporated the remaining variables and interacted with the same level-one predictor. Statistically significant level-two predictors of nonresponse were part-times students, athletes, and students who were 25 years and younger (traditional college age). However, students living on-campus and transfer students were more likely to respond than their counterparts. These differences were consistent with previous analyses in this chapter (Tables 4.4 and 4.5). When a student took a class in the same department as their major, the statistics were no longer significant. Traditional age students, athletes, and part-time students did not maintain significant differences from the average likelihood of response. In Table 4.5 (models for hypothesis one variables), each race category was statistically significant and less likely to respond than the reference group (excluding those who did not report their race). Interactions with the course location predictor also affected these groups. They were not significantly different from the mean likelihood of response when a course was in the same department as their major.

The fully controlled model (Appendix O) did not vary considerably from the uncontrolled model. This model produced similar statistics, especially in the no-interaction models. Some differences were evident with transfer students, the other race category, and part-time students. In the fully controlled models, transfer students increased their significance (ρ-values), while the statistics for part-time student and the "other race" group were no longer significant.

Hypothesis Five

Survey salience will increase the likelihood of response. I analyzed measures of academic environment of the student's major, academic environment of the course, and grades. Several of these influences tended to increase or decrease the salience of the SET, influencing the likelihood of response. In hypothesis four, I already modeled a form of salience with academic environment and course location (in or outside of the department that houses the student's major) (Table 4.10). The results from that model are also appropriate here, because a course in a student's academic environment was one of the most salient attributes measurable.

Grades are also a measure of salience, and interactions with academic environment are in the table below (Table 4.12). Because of error messages received due to high correlations and large file size, variables required adjustments in the full models. An unconditional model for grades and environment, without these adjustments is in Appendix Q. New variables combined the grades of D and F, the no course grade (NG) category remained, and A, B, C, and S were the reference group. Figure 4.2 showed that these four groups did not significantly differ from the average response rate for all students, but the D, F, and NG grades responded at a rate below 40%. Appendix P provides the fully conditional model statistics and all other variables.

Table 4.12. **Salience of grades and environment: Uncontrolled and fully controlled models.**

Variable		***Grades and Environment***		***Controlled No Interactions***		***Fully Controlled***	
		Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Level One							
Grade	(Intercepts)						
	D and F	-0.261***	0.770	-0.189***	0.828	-0.261***	0.770
	No Grade (NG)	-0.453***	0.636	-0.456***	0.634	-0.453***	0.636
Level Two	(Intercepts)	-0.059***	0.942	-0.069***	0.933	-0.073***	0.930
Academic Environ.							
(*Social = ref.)*							
	Artistic	0.038	1.039	0.052	1.053	0.051	1.052
	Conventional	0.025	1.025	0.077	1.080	0.088	1.092
	Enterprising	-0.173***	0.841	-0.039	0.962	-0.049	0.952
	Investigative	-0.058	0.943	0.022	1.023	0.031	1.032
	Realistic	0.055	1.057	0.204***	1.226	0.213***	1.238
	FYC/Transitions	0.001	1.001	0.130*	1.139	0.131**	1.140
Interactions (* D & F)		-0.275***	0.759	-----	-----	-0.254***	0.776
	Artistic	-0.229*	0.795	-----	-----	-0.158	0.854
	Conventional	0.140	1.151	-----	-----	0.203	1.225
	Enterprising	-0.082	0.921	-----	-----	-0.072	0.931
	Investigative	0.057	1.059	-----	-----	0.056	1.057
	Realistic	0.112	1.119	-----	-----	0.107	1.112
	FYC/Transitions	0.058	1.060	-----	-----	0.008	1.008
				-----	-----		
Interactions (* NG)		-0.488***	0.614	-----	-----	-0.501***	0.606
	Artistic	0.031	1.032	-----	-----	0.076	1.079
	Conventional	0.084	1.088	-----	-----	0.080	1.083
	Enterprising	-0.025	0.976	-----	-----	-0.033	0.968
	Investigative	0.246***	1.279	-----	-----	0.252***	1.288
	Realistic	0.190**	1.210	-----	-----	0.196**	1.218
	FYC/Transitions	0.048	1.049	-----	-----	0.046	1.047
	Tau	6.404		6.173		6.190	
	Intercept Reliability	0.751		0.735		0.734	

*** p<0.001, **p<0.01, *p<0.05

Overall, no grades, Ds, and Fs were significantly less likely to respond than the reference group (As, Bs, Cs, Satisfactory). The outcomes of both models produced very small differences when comparing them (uncontrolled and controlled) with three

exceptions. Enterprising was significantly less likely to respond in the uncontrolled model without interactions, but was without significance in the fully controlled model. Conversely, in the same model realistic and first-year groups were statistically insignificant in the uncontrolled model, but produced significant statistics in the full model.

SETs for courses in which students did not receive a grade were about 0.64 times as likely to see submission. These odds ratios did not change much for interaction intercepts for both the unconditional model and the fully conditional model (submission was about 0.61 times as likely). Two groups in the no grade interactions were significantly different from the average probability of response. Investigative (OR = 1.28) and realistic (OR = 1.21) types were more likely to respond.

Answers to the Research Questions

1. *To what extent was the likelihood of response to SETs related to student-level measures, such as demographics or academic performance?*

Several level-two variables emerged as predictors of SET response. All females, race, and age were statistically significant. Specifically, females respond more than males, traditional-aged students were less likely to respond than older students, and African-American and Asian students were less likely to respond than white/Caucasian students. In conclusion, males, African-Americans, and traditional-aged students were the most significant demographic predictors of nonresponse for this population.

The student's semester GPA was very significant in predicting nonresponse. When compared to the mean response rate, GPAs below 3.0 highly predicted nonresponse. Grades, especially Ds, Fs, and No Grades, were also predictive of nonresponse at level-one, as submission for an SET associated with such a grade was 0.64 times as likely as a grade of A, B, or C. Other student level-characteristics predicting nonresponse were class rank (sophomores and juniors), realistic majors, and transfer students.

2. *To what extent was the likelihood of response related to personality and academic environment as described by Holland?*

Realistic majors displayed a propensity to respond to SETs. However, course location increased the likelihood of response for all major types. Overall, course location was very significant in predicting SET participation. SETs for courses in the same department as the student's major influenced level-two variables to the extent that a majority of characteristics did not differ from the mean with this interaction. Sophomores and African Americans, previously very unlikely to respond, were more likely to respond if the course was in the department of their major.

3. *To what extent did course salience increase the likelihood of response? Did the relationship vary significantly between students? Could student-level measures explain this variation?*

Course salience increased the likelihood of response significantly. Submission of SETs associated with grades of Ds, Fs, and No Grades was less likely than As, Bs, and Cs. Again, the course location variable (in the same department as the student's major) was a predictor when interacting with level-two variables. High grades and a more familiar academic environment increased the likelihood of response.

Course salience did not greatly vary between students with the interactions of level-two variables. For most students, regardless of age, semester GPA, number of credit hours, gender, etc., they were just as likely to respond with these interactions. Course location only significantly increased the propensity to respond for African Americans, courses in the student's academic environment, and sophomores; the likelihood of response decreased for transfer students with this interaction.

4. *To what extent did survey fatigue predict response rates? Did response rates correlate to opportunity costs? To what extent did culture vary response rates?*

Characteristics of survey fatigue for this study, SETs to complete and class rank according to credit hours, were significant in predicting nonresponse. However, seniors did not differ from first-year students. Survey fatigue in this generation of web-based surveys is difficult to measure, due to their virtually omnipresent existence. More measures of survey fatigue on and off campus would be helpful to future studies.

Opportunity costs measures were significant for most variables. Part-time students, those with more SETs to complete, commuters, athletes, sophomores, and juniors were less likely to respond. Seniors were just as likely as first-year students to respond.

Culture of the academic environment, both of the course and the student, varied response rates. However, while academic environment increased the propensity of SET submission, most culture variables available to this study did not behave as expected. Realistic types were more likely to respond to any SET, while artistic students were more likely to submit an SET for a course in their department and social majors were less likely to submit an SET for a course in their department.

Other culture variables did not vary response rates as expected. For example, social exchange theory proposes that survey participation increases for participants who have closer relationships to their environment or community. On-campus students did tend to respond more than commuters, but social exchange theory did not work for all measures in this study. Transfer students and older students were actually more likely to respond while the amount of NC State credit hours, part-time students, and residents of North Carolina were not significant predictors. Conversely, athletes were less likely to respond. However, academic environment accounted for the significance.

Summary

Several level-two variables exhibited statistically significant differences, and their interactions with level-one variables many times increased or decreased their probability of response. Even for predictors of consistently high significance (e.g. housing, race),

certain level-one predictors (e.g. course location) tended to change those outcomes in the likelihood of response. The next and final chapter explains the results in the context of the hypotheses, theories of nonresponse, and research questions. Additionally, I present implications and practical applications based on these findings, theories, and past research.

CHAPTER FIVE: IMPLICATIONS AND CONCLUSIONS

Many colleges and universities now administer end-of-course evaluations online instead of face-to-face, allowing students to complete their evaluations at a self-determined time. Like other online surveys, many institutions consequently have seen response rates decrease, which increases the potential for nonresponse error (Dillman et al., 2009; Fowler, 2009; Groves et al., 2009). Unlike most other online surveys, these results can potentially have high stakes, because administrators consider responses when making personnel and curricular decisions. To lower the risk for error and increase response rates, institutions must understand nonresponse on their campuses. That was the basis of this study. By applying nonresponse theories and previous research, this study analyzed undergraduate participation in course evaluations, also called student evaluations of teaching (SETs), for the Fall 2009 semester at North Carolina State University (NC State).

Nonresponse Bias: Error in SETs

Before the advent of web-based surveys, instructors administered a paper-based SET during class time, and coverage error was more prevalent. Nonresponse error, commonly associated with web-based surveys since participants typically respond in their own time, is often reported in the form of bias. Nonresponse bias occurs when patterns of nonresponse exist across particular groups. This study identified patterns in participation, that is, groups or populations who were typical nonrespondents. On this campus, there is evidence of SET nonresponse bias, as certain types of students did not respond at the

same rate as their counterparts. Females, students housed on campus, first-year and senior class ranks, realistic majors, older students, and races other than African American and Asian categories were significantly more likely to respond. Respondents typically had higher GPAs, and submitted SETs typically had an A, B, or C course grade attached. Realizing that valid responses can also come from students with low or no grades, males, juniors, sophomores, commuters, and traditional-aged students, their responses are important to data quality and error-reduction.

Furthermore, nonresponse bias may double-bias results. Biases that emerged from previous research on SET ratings many times also match biases associated with nonresponse in this and other studies. For example, this and other studies have found that students with higher grades are more likely to respond, but previous SET research also links higher ratings to higher grades. Not only are students with higher grades typically awarding higher ratings, but they are also the ones who are more likely to respond. Can instructors make appropriate changes to their courses if lower-achieving students are not responding? Reviewers of the ratings should always consider the added bias that nonresponse introduces. Knowing characteristics consistent with nonresponse allows survey administrators to direct their efforts to reach these students and reduce nonresponse error.

Theory, Previous Research, and Implications

At times, the results reinforced nonresponse theories and previous survey research; at others, contradictions were evident. Because hypotheses directly reflect theories and previous research, they serve as a framework below to present the

relationships among outcomes, theory, and research. Following this analysis, I address more specific implications, considerations for future research, and practical applications for institutions and researchers.

Hypothesis One: Demographics

Females, white students, and older students will be more likely to respond to SETs than their peers. Evidence from this study supports this hypothesis. Gender and race were significant factors when predicting the likelihood of response, consistent with previous studies on survey nonresponse (Avery et al., 2006; Dey, 1997; Fidelman, 2007; Johnson et al., 2002; Jones, 2009; Lepkowski & Couper, 2002; Moore & Tarnai, 2002; Porter & Umbach, 2006a; Porter & Whitcomb, 2005; Sax et al., 2003). Age was also significant, as nontraditional students were more likely to respond.

Studies of nonresponse on college campuses need to incorporate age more consistently. Few studies have addressed this characteristic, but those that do have also found that students of college-age were less likely to respond than older students (Kaplowitz et al., 2004; Sax et al., 2003, 2008). Postsecondary populations are typically between the ages of 18-24, but with current high unemployment rates, undergraduates over 25 years of age are a growing population.

Implications.

Demographics inform NC State and other institutions about nonrespondents. With this information colleges, departments, and other organizations are able to target those who do not respond. Could an SET request from a male influence other males to respond more? Could diverse organizations on campus encourage racially diverse populations to

submit SETs? With collaboration and effective communication, a variety of efforts can address nonresponse and influence nonrespondents to become respondents. Moreover, those who use SET results should be aware of these biases.

A double-bias, as previously addressed, may exist for gender and race especially. For example, when compared to males, some studies have found that females tend to give higher ratings for their instructors (Basow, 1995; Cramer & Alexitch, 2000; Tatro, 1995), and they are more likely to submit their SETs. At a campus that is almost 57% male, how does this effect data quality? Are decisions made with these data accurate? For data that are more representative of the population, targeted efforts to raise response rates for these groups would potentially increase data quality.

Finally, data analysts could apply nonresponse weights to those who have responded. Post-stratification is a common nonresponse weighting procedure. Weighting may increase the precision of the estimates, as it would help represent previously underrepresented groups. However because weights rely upon those that did answer, weighting may not produce accurate results or accurately portray groups for whom the weighting is intended to benefit.

Hypothesis Two: Academic Performance

High academic performance will increase the likelihood of SET response. Results from this study support this hypothesis. A higher semester GPA and a higher grade in the course were positively related to the likelihood of response, which is a consistent theme for postsecondary surveys (Avery et al., 2006; Dey, 1997; Fidelman, 2007; Porter &

Umbach, 2006a; Sax et al., 2003, 2008). These results lead to multiple questions for institutions and the SET process.

Why are higher-achieving students consistently more likely to respond? In this study, perhaps they felt confident that they had the intellectual ability to evaluate the course at a meaningful and variable level. It is also possible that students perceive non-credit and pass/fail courses as less important and less deserving of their attention. It is possible that valid reasons for low grades may also have prevented response, and a consequence of unfortunate life circumstances may be a low grade. If student did not attend classes or complete assignments, feelings of inadequacy to properly evaluate the course may have influenced nonresponse.

Failing grades were not the only grade-based influences on nonresponse; no grades were also significant. No grades outcomes refer to ungraded courses, audit grades, withdrawals, and incompletes, where submission of the SET was 0.64 times as likely. Are evaluations valid for these ungraded courses and students? Should SETs be requested of these students?

Implications.

There are several considerations helpful to campuses in addressing nonresponse. Labs comprise a large majority of ungraded courses. However, the core question set (Appendix A) already contains lab questions. Students have already provided information for the lab in the graded course SET, so do they need to complete another evaluation stating the same thing? Because course codes for labs either vary by college or have no identifier at all, there is no way to identify labs and make this conclusion. However,

identified labs could have a separate instrument to prevent multiple evaluations, which may indicate an issue with data-quality, coverage error, and over-sampling.

Are instructors receiving high-quality feedback about instruction only from students with the best grades? Do they have good data with which they can make meaningful changes to their teaching strategies? Institutions need to explore the added bias associated with nonresponse. Not only do students with higher grades tend to rate more favorably their instructors (Cohen, 1981; Costin, 1978; Crumbley & Reichelt, 2009; Isely & Singh, 2005; Marsh, 2007; McKeachie 1969), but also these are the students who are more likely to respond at all. This double-bias, evident in respondents and how they rate their instructors, introduces an increased risk for nonresponse error.

Additional questions arise about withdrawals, incompletes, and auditing students. Typically they have not fully participated in the class and completed all requirements. When several questions in the instrument address course readings, assignments, and instructor feedback (Appendix A), these students may have felt as if they could not properly evaluate the course. Moreover, should they be able to evaluate the course? Like ungraded lab courses associated with graded courses, perhaps different instruments may be appropriate for these students and courses. This would help decrease the risk for coverage and nonresponse errors, survey fatigue, and the general frustration that students feel when evaluating a course multiple times. High performance may also have a relationship with topic salience, addressed in hypothesis five.

Hypothesis Three: Opportunity Costs and Survey Fatigue

Opportunity costs and survey fatigue will decrease the probability of response. These two theories both used class rank as predictors, a factor significant in SET response. Class rank has been influential in participation in other studies (Davis et al., 2007; Fidelman, 2007; Porter & Umbach, 2006a; Porter & Whitcomb, 2005). However, the ways in which these factors influenced participation are not consistent across this and other studies. For example, upper classmen are typically nonrespondents (Laurie et al., 1999; McGourty et al., 2002b), but at this institution, seniors were not significantly different from first-year students, and sophomores and juniors were less likely to respond than first-year students. Sophomores and juniors may have felt over-surveyed after experiencing an abundance of first-year surveys. In high school and before, students rarely have opportunities to provide their opinions. Once in college, they have many opportunities to communicate feedback. However, with survey requests originating from student affairs, campus dining, campus housing, national agencies, sophomore surveys, program surveys, department and college surveys, what may once have been an opportunity to finally voice an opinion may have become cumbersome after year one. These two classes may have had to focus on program requirements and entrance into their majors as sophomores and juniors, indicative of opportunity costs.

As seniors, students did not respond differently than first-year students, possibly for several reasons. First, perhaps they had previously followed the same pattern of nonresponse in their sophomore and junior semesters. As seniors, they could have felt that they had not “done enough” and needed to make up for previous terms (Laurie et al.,

1999; Porter et al., 2004). Second, perhaps they felt that they were now capable enough to analyze and provide good feedback for instruction. Third, considering social exchange theory (Blau, 1964; Dillman, 1978), seniors may have taken advantage of their final change to "give back" to the institution. Seniors' final courses are typically in their major, which was a strong interaction variable (see the next hypothesis).

Survey Fatigue.

Modern society has become increasingly dependent upon the online environment. Technology has overhauled survey design and administration, allowing more surveys to reach more people. Consequently, evidence of survey fatigue is on the rise.

In this study, the number of SETs administered to the student was a statistically significant predictor of participation; the likelihood of response decreased with the number of SETs administered to the student (see Figure 4.5 where SET submission rates decreased with 10 or more SETs administered to the student). The number of SETs administered is not a variable often seen in nonresponse research. However, with technology increasing survey requests, this factor may continue to be indicative of survey fatigue on campuses in future studies.

With an increasing number of survey requests sent to students' email inboxes, those that arrive last might be most prone to survey fatigue, consequently contributing to lower response rates (Groves et al., 2004; Porter et al., 2004). Many campus offices and departments administer surveys to students throughout the semester, but SETs are the last ones students receive. There is also the potential for off-campus surveys to contribute to

survey fatigue, and to provide a fuller context for survey fatigue future research should consider total survey requests in postsecondary student populations.

Opportunity Costs.

At NC State, there is evidence of nonresponse attributed to busy schedule or the lack of opportunities to complete the administered survey (Dillman et al., 2002; Groves et al., 2004). From an opportunity-costs perspective, perhaps sophomores and juniors were not as likely to respond because these students are beginning to enter their major coursework. They may prioritize applications to majors or programs, completing requirements for entrance into majors or programs, or becoming accustomed to upper-level coursework.

Commuters were less likely to respond than on-campus residents, and athletes were less likely to respond than non-athletes. Commuting students take time to drive to campus, probably have jobs, and may have other responsibilities than on-campus students would have. Students living on campus had more opportunities (e.g. on-campus computers, free internet) and constant exposure to posters and advertisements reminding them to complete these surveys. Athletes also have more responsibilities to their team, tutors, coaches, and practices. Part-time students did not differ from full-time students in their likelihood to submit SETs.

Implications.

Evidence from this study supported this hypothesis, and several implications emerge from these outcomes. First, more information would validate results and inform continuing research (Groves & Couper, 1998; Groves et al., 2004; Porter et al., 2004).

Several measures of survey fatigue and opportunity costs are missing from this study. For example, the total number of survey requests an individual receives and each individual's priorities and situations are largely unknown, making definitive results based on these two theories of nonresponse difficult to assess. Other studies should consider the incorporation of variables such as these that are indicative of nonresponse associated with survey fatigue and opportunity costs.

Second, campuses can limit survey requests to the student body and centralize the survey process in an institutional research office. How many other surveys are students receiving from the university and its affiliated departments and programs? Are they also unresponsive to those after their first year but not in their senior year? Most campuses never know how many surveys students receive from campus offices, nor do they know who sends them. Because of these issues, centralization and control of the survey administration processes on campus is an increasing reaction to declining response rates. Some campuses even sample their students for external or internal survey requests, and provide a different sample for the next survey request.

Third, institutions can target their unresponsive populations. Coaches can provide time for their athletes to complete SETs, and commuter-computer labs can provide internet access to off-campus residents who need to complete their SETs. Fourth, institutions that have already implemented similar measures to address nonresponse can communicate their results. What differences do they notice and are their plans working? This knowledge also would be helpful to test these two theories as we increase our understanding of nonresponse in an online environment at college campuses. Further

research would also provide more information for other institutions. It also may be necessary that campuses consider broad factors of students' contexts, both on and off campus, to appropriately address nonresponse.

Finally, SET administrators could consider sampling students for SETs and any other online survey requested of students. Because of the evidence of survey fatigue, students could be limited to no more than 10 SETs to complete (see Figure 4.5), by sampling or by their own choice of SETs to complete. Sampling may not include the perspectives and opinions of all students, and this method of surveying should be thoroughly explored. Typically survey methodologists encourage population sampling when possible, but alternatives may be necessary when addressing survey fatigue. Changing the way in which labs are evaluated (see implications for hypothesis two), helping to eliminate oversampling, may also reduce survey fatigue and increase response rates.

Hypothesis Four: Culture

The probability of response will increase for students with stronger cultural and communal ties to the campus. Consciously and subconsciously culture influences decisions, including the decision to complete a survey. Evidence to support this hypothesis emerged from this study. For postsecondary students at NC State in the 2009 fall semester, factors associated with campus, academic, and personal culture increased or decreased the likelihood of SET submission (Blau, 1964; Dillman et al., 2002; Emerson, 1976; Groves et al., 1992).

Academic Environment.

Realistic majors had a propensity to respond greater than social, artistic, conventional, enterprising, or investigative majors (see Rosen et al., 1997; see also Appendix I). Because realistic majors, according to Holland (1973), are practical and avoid social interactions, they may feel more comfortable completing SETs in an online environment instead of a classroom. Realistic environments that house these majors may appreciate and encourage participation in online surveys, since they do not have to interact with other students while completing it (Smart et al., 2000).

These outcomes are different from those of previous studies. Porter and Whitcomb (2005) found a positive relationship between survey participation and two of the six categories, social and investigative. Other studies have found increased response rates for extroverts and social students (Sax et al., 2003; Yu et al., 2007). Realistic majors are not social or extroverts; on the contrary, they avoid social interactions. Discrepancies may exist because the incidence of student types varies by institution or semester or geography, or perhaps the variance resides in secondary and tertiary characteristics of the environment (Rosen et al., 1997; Smart et al., 2000). Nevertheless, environmental and community factors, as attributed to course location and academic major, require additional research in survey participation relevant to postsecondary institutions.

For academic environment, a course in the same department as the student's majors was more likely to have a submitted SET, as predicted by the theory of social exchange. Because these students had strong ties to their academic community, they probably felt invested in the group's interests and its betterment. Students made the

connection with their environment, and exhibited helping behaviors by participating (Johnson et al., 2002; Porter & Whitcomb, 2007). Perhaps they felt a sense of duty to reciprocate for the services and support provided by their department (Blau, 1964; Clarksberg et al., 2008; Porter & Umbach, 2006a). Additionally, they may have felt more knowledgeable about that subject in their own environment and academic construct, and this may have contributed to the increased likelihood of response.

Additional Measures of Culture.

Considering the theories of social exchange and culture, predictions may assume nonrespondents to be non-residents, students who have not been at NC State as long, transfer students, non-athletes, commuters, older students, and part-time students. Of these predictions, only commuters were significant factors of nonresponse. Conversely, transfers were actually significant predictors of response, while part-time status, NC State credit hours, and residency were not significant. Traditional-aged students and athletes were significantly less likely to respond.

Can opportunity costs explain these outcomes? Commuters have other responsibilities. Athletes may have been too busy and felt that they have already contributed to the campus culture by their representation in NC State athletics. However, several groups are exempt from opportunity costs reasoning, because older, transfer, and part-time students also tend to have more responsibilities. Instead they were not statistically significant (part-time) or were actually more likely to respond.

Lack of roots in the campus community should have also influenced transfer students' response rates negatively. The number of credit hours earned at the institution

should have influenced higher response rates according to social exchange. However, completing SETs may exhibit these students' attempts to try harder to reach out and be involved in their new community. Higher response rates for older students aligned with previous research (Kaplowitz et al., 2004; Sax et al., 2003, 2008). Though they may not have time to become more engaged in NC State's culture, they wanted to contribute in this way by completing their SETs.

Course Location.

Most previously significant predictors were no longer significant with this interaction. However, two exceptions were athletes and transfers who remained significant, but with alterations. Athletes became positive predictors of response, meaning that their environment significantly influenced SET submission. When they were in a familiar setting or environment, they were more likely to respond. Transfer students became negative predictors of response, a phenomenon unseen in any other models where they were consistently more likely to respond. Even when interacted with a grade of D or F, transfer students were never significantly less likely to respond. However, with this interaction, transfer status was the only variable that was a statistically significant predictor of nonresponse. Why were they less likely to respond when the SET was for a course in the same department as the student's major? Perhaps their major courses are much more difficult than their previous college or general education courses, or perhaps they feel out of place because they have not been in that major as long as their peers. Further research of transfer students within their majors may be necessary to address in continued analyses and studies of survey participation.

Second, the concept and theory of *minority group opposition*, evidenced through a lower rate of submission of SETs by non-white students, may have been a factor in nonresponse (Johnson et al., 2002). This lower level of response may have represented feelings of mistrust and resentment towards the majority administration. But these groups of students did not appear to feel this "opposition" if the course was in their major department. No differences among groups emerged with the course location variable.

Third, with course location, effects of response also differed with this interaction of class rank. Juniors, first-years, and seniors were not significantly different in their likelihood of response. However, sophomores, who in other models were significantly less likely to respond than first-years, became more likely to respond, though the response rate did not reach statistical significance. Perhaps sophomores and juniors in this case contribute to social exchange – they are beginning to take more courses in their majors, taking fewer general education courses, and want to contribute to their academic "family" (Blau, 1964; Dillman et al., 2002; Emerson, 1976; Groves et al., 1992). Social exchange and academic culture also may be the reasons for seniors to be consistently as likely as first-year students to complete SETs. Seniors may have taken advantage of their final chance to "give back" to the institution. Moreover, seniors' final courses are typically in their major.

Implications.

These results have important implications for campuses to combat nonresponse in these online evaluations. First, students' major departments can communicate the need to

complete SETs. Realistic majors are more likely to respond, but the other five types could communicate SET completion as suitable and salient to these students as well (see the final hypothesis).

Second, institutions should consider using social exchange to their advantage. Organizational structure, particularly institutional culture or a symbolic focus (Bolman & Deal, 2008), can increase feelings of belonging and promote survey completion based on social exchange concepts. To create a culture of survey response on campus, institutions can consider promoting the ritual aspects of SETs, integrating stories about SETs making a difference in instruction and curricula, and creating a sense of belonging and importance for all members (see also Dillman et al., 2002; Groves et al., 2004; Hofstede, 1981).

Third, academic disciplines and other organizations or offices associated with populations of nonrespondents can also utilize social exchange. African American and other organizations can encourage students of color to voice their opinions and indicate that their opinions are important. Commuter support services, athletic teams, sophomores and juniors, and academic support services for students with low grades can similarly encourage their target support groups. Academic environment of the course appears to have an influence on participation. Students in most disciplines (Figure 4.6) were more likely to respond to SETs for courses in their own majors. To communicate that all SETs are important, the onus may be on these academic disciplines. Moreover, if students feel uncomfortable rating a course outside their major, their academic discipline needs to convey confidence. They could provide information, workshops, or an orientation to the

ClassEval process to encourage response, because students have the ability to critique and comment on any course.

Lastly, survey research on postsecondary students must continue to address academic, campus, and personal culture in survey participation. Culture influences almost any decision (Hofstede, 1981). Because culture places people in environments and communities where they feel safe and comfortable, it may be an integral component in reaching nonrespondents via communities. While social exchange was influential in this study, course location is rarely part of SET research, if at all. Further studies should include this factor, since academic environments can encourage students to become respondents.

Hypothesis Five: Salience

Survey salience will increase the likelihood of response. Past research has demonstrated that survey response is more likely if the individual sees a survey as beneficial, interesting, and appealing (Groves et al., 2004; Groves et al., 2006; Groves & Peytcheva, 2008; Groves, Presser, et al., 2004; Heberlein & Baumgartner, 1978; Remedios & Lieberman, 2008). Unfortunately, few if any studies on postsecondary student surveys have addressed topic salience as measured by the course location. However, grades or expected grades were consistent with previous research.

In the previous hypothesis, type of major and course location were measures of academic culture or environment, but these are also measures of salience because the student's choice of major represents an area of interest. First, outcomes revealed that realistic majors were the only statistically significant predictor of response. This indicates

that SETs were already salient to students in realistic disciplines. Second, as previously discussed course location was influential. If the course was in the same department as the academic major of the student, it was more likely an area of interest. Therefore, SET submission was more likely. The probability of response dramatically decreased for social majors if the course matched their major's department (Figure 4.6). What about these courses decreases salience for social majors but increases salience for others? This is an anomaly that needs further research and closer examination.

Another salience measure was grades. Naturally, we avoid reminders of our failures and shortcomings. An SET would be such a reminder if the grade earned for the course were an F or U. Similar to past studies (Avery et al., 2006; Clarksberg et al., 2008; Dey, 1997; Fidelman, 2007; Porter & Umbach, 2006a; Sax et al., 2003; Sax et al., 2008), students who could associate that evaluation with something good, in this case a high grade, survey participation was more likely than those who received an A, B, or C, even with interactions. Failing grades were not the only non-salient category; no grades were also significant but are largely absent from SET research. As discussed in hypothesis two, no grades are ungraded courses, audit grades, withdrawals, and incompletes. Whether it was an ungraded course, such as a lab, or the student did not complete all the requirements for credit, issues related to coverage error, salience, and survey fatigue may influence nonresponse.

Implications.

It is possible to increase the salience of SETs, potentially increasing response rates. First, this can be accomplished through the values and priorities communicated by

the type of the student's major. While realistic majors were more likely to respond, the other five major types could establish reasons why SETs are also salient to their disciplines. Using Holland's descriptions of these six types, these disciplines could make their students aware of how salient SETs really are to their personalities. There are several potential arguments. By completing SETs, social majors are contributing to the betterment of the community and participating in this social activity (Dillman et al., 2002). Investigative majors would be contributing to data collection and research. Conventional majors could see the order and organization of the SET process as attractive. If they responded to their SETs, enterprising majors would be providing feedback to persuade their instructors to change or not change. Artistic majors could interpret this process as a way they can express what they feel and how they perceive their courses. Any environment could build a case for SET response, but they must communicate these values to their students.

Second, ungraded courses may not be salient, because students may not see them as important. They may have a good reason. Students have already provided information for the labs in the SET for the graded course; and audits, incompletes, and withdrawals have not completed all course requirements. A non-grade introduces several questions for SET administrators. Are typical evaluation instruments valid for these ungraded courses and students? Do these students have enough information to evaluate well that course and instructor? If the ungraded course is a lab associated with a graded course, does this introduce coverage error in the form of oversampling?

Finally, more variables would help to determine salience. Do they really like their major or are they in that discipline for wealth or because of parental influence? Did they fail a salient course because of unfortunate circumstances? Each individual's personal tastes and interests may be difficult to obtain and measure. Additionally, salience can change; perhaps someone switches from a realistic discipline to an enterprising one. It is becoming increasingly evident that more research into the salience of survey participation, especially for college students and SETs, is necessary.

Now What? Extending the Conversation

The section on nonresponse bias and each hypothesis have already addressed implications of SET nonresponse. This next section extends those discussions. For institutions similarly experiencing nonresponse on their campuses, what can they do throughout to prevent nonresponse? Outcomes from this study, nonresponse theory, and previous research point to four main implications. Background efforts, survey design, methods of survey administration, and emerging institutional practices may be effective to combat nonresponse in college student populations.

Background Efforts

Overall, SETs are important to the institutional culture. Smart et al. (2000) and others would recommend a deeper examination of the institution, addressing the cultures within academic disciplines. "If we stick to the naïve assumption that because they look just like us, they also think just like us, [then] our joint efforts will not get very far" (Hofstede, 1981, p. 35). This and other studies found exactly that; even though all ClassEval participants attend NC State, they do not necessarily respond at a similar rate.

How could institutional survey administrators appeal to these academic environmental influences? Anthropologically, we crave belonging and the establishment of communities or structures with which to associate (Lévi-Strauss, 1967), and both faculty and students need to belong. Considering the significance of students' major types and course location, institutions can begin to establish a culture that values SETs in two ways. First, they should communicate and reinforce the importance of all SETs to instructors who influence the environment and to students who acculturate to that environment (Etzioni, 1975; Hofstede, 1981). Second, within or outside the academic environment, providing skills and knowledge can orient students to the SET process. These two acts of enculturation create institutional and academic environments wherein students can feel comfortable and able to provide valuable feedback for SETs inside or outside their major.

Within the Institution: Communicating Values.

The influences of culture on decision-making processes are strong, but cultural significance is weak without communication. Increased communication about SETs from the institution to faculty and students can create a sense of belonging. Social exchange may then decrease the effects of survey fatigue and opportunity costs on response rates.

At NC State, communication from UPA travels to students and faculty mostly through email. When sending emails about surveys, Dillman et al. (2009) recommend that that they be concise and straightforward. Also, multiple emails and reminders are integral components to reduce nonresponse (Dillman et al., 2009; Groves et al., 2009; Nulty, 2008). Only one email provided ClassEval information to faculty (Appendix E).

The length of this email was formidable and needed to be more concise. Division of this email into three or more emails would increase their attention. Length of communication is important to maintain readership; especially considering that faculty attitudes and efforts are key to obtaining higher response rates (Ballantyne, 2003; Dillman, 1978; Dillman et al., 2002, 2009; Dommeyer et al., 2004; Groves et al., 1992). Multiple emails would remind instructors to remind their students and convey that SETs are valuable and meaningful. In these emails and in other forums, institutions should communicate to faculty that encouraging SET completion is a sign of good teaching. Higher response rates are more likely to provide quality feedback for instructional evaluation, and good teachers seek to improve instruction. Consistent and concise communication based on researched survey practices may lead to instructors' valuing SETs and persuading others in their environment to do the same. Administrators, faculty, students, and other stakeholders could reinforce institutional and academic cultures; positive and constructive communication among these groups is essential.

Student emails may or may not have provided enough information about ClassEval. Appendices C and D indicate that frequent email reminders from the UPA ClassEval email address encouraged SET participation, consistent with survey methodology and efforts to combat nonresponse (Dillman et al., 2009; Groves et al., 2009; Nulty, 2008). However, more precise and concise information help students better understand why, where, and how the process works (Dommeyer et al., 2004). Email requests could include or provide links to more information, but communication of this information should happen before the SET request.

Faculty orientation or workshops can inform instructors, explaining the whys and hows and values of the SET process. Similarly, student orientation may also be effective to teach students how to complete the instrument and provide adequate information. Even if students did not excel in a course or if it is a course in which they have little interest or knowledge, they should have confidence to evaluate it. Membership to institutional culture is more meaningful if one can contribute and know that contributions are significant and provide helpful feedback.

Should email requests be the only form of communication upon which the institution relies? Most courses do not meet in an online-only environment, and NC State is not structured as an internet-only or internet-based institution. Below are suggestions for institutions and SET administrators, each one associated with a theory of survey participation whose measures were significant predictors in this study. These suggestions are additional communicative efforts to combat nonresponse that institutions should consider providing to their participants:

1. A website (FAQ), advertisement in the student newspaper, or speakers in classrooms and workshops could address:
    a. *Culture.* SET procedures, including confidentiality (Anderson et al., 2005; Nulty, 2008; Quinn, 2002).
    b. *Culture and survey fatigue.* This is not just another survey. So, for what purposes does the university collect SETs (Marlin, 1987; Zúñiga, 2004)?

c. *Minority group oppression*. Is there another person I can ask questions about SETs? (Johnson et al., 2002).

2. *Opportunity costs*. Information about computer labs dedicated to SETs during the data collection period, or information about open computers where students could complete their SETs.
3. *Academic environment and culture.* Further evidence that their opinions are valuable and useful for all courses. Perhaps only one professor's opinion is not enough, since the story repeats (Appendices B and D).
4. *Salience and social exchange.* Demonstrate to students how to respond to SETs and give constructive feedback (Dommeyer et al., 2004; Zúñiga, 2004), especially in cases where they may not be familiar with an online survey environment (Quinn, 2002).

Within the Environment: Providing Knowledge and Skills.

Attempts to establish a culture that values SETs do not stop at the institutional level, especially for large universities and colleges. Faculty members have the ability to persuade members within their academic culture and environment. They have the power to reinforce organizational values and requirements (Etzioni, 1975; Hofstede, 1981). Faculty members are in a position to perpetuate the ideals of the academic discipline and enculturate their students, but their interest in exerting this influence through is declining (Smart et al., 2000). Therefore, faculty buy-in of culture and social exchange theories would be crucial to establishing an academic culture of response.

With this information, students and faculty may begin understanding the importance of the SET process and want to contribute to their campus. Students need to hear that they can contribute to this process, no matter what their subject expertise may be. Organizations on campus, academic or non-academic, to which these students belong can have a role in addressing nonresponse if their group members have been identified as less likely to respond. Encouragement from several organizations, environments, and groups on campus may boost their confidence and SET response rates.

SET Instrument Design

Lack of salience and factors related to opportunity costs may take the form of inapplicable questions or too many questions. The current ClassEval instrument is three years old and contains 19 questions, but instructors, colleges, and departments can add additional questions. It is likely that some SETs are much longer than others. Moreover, with many online surveys now containing fewer questions and less content, reconsideration of the instrument's length may be necessary to maintain salience (Quinn, 2002). Lab questions (numbers 15-18) may also be confusing to students not participating in such a course. Inclusion and exclusion of lab questions may also require further consideration, because if the lab has links to another course, the student may be evaluating the same course at least twice. If there is more than one instructor, each instructor also requires an evaluation. Essentially, the student could be answering the same questions for the same instructor and the same class in multiple SETs.

To address this, modified SETs are appropriate for courses with labs and for the labs only. The “Tailored Design Method” (Dillman et al., 2009) advocates modifications

of survey administration and data collection to fit the participants' needs and preferences, reducing the risk for error and increasing response rates. However, to appropriate tailor the design of SETs, lab courses must have proper identification via course codes or a check box when creating a course. Modified SETs for auditing students, withdrawals, or incompletes also require consideration. Are these students able to evaluation course assignments, instructor feedback, and course readings? "Not applicable" options are always available, but are problematic (Dillman et al., 2009). If well made, shorter surveys typically have higher response rates and can still have meaningful results (Dillman et al., 2009; Heberlein & Baumgartner, 1978).

SET Administration and Data Collection

Social exchange is present in all parts of the survey process. For administration and data collection, academic disciplines may be able to contribute to survey completion. SETs requests could originate from the students' major department. The academic community to which the student belongs also could communicate encouragement throughout the ClassEval data collection period.

Opportunity costs may also demand that online SET administration evolve. Is it necessary to relegate SET completion at the students' leisure only? Perhaps not. Like paper evaluations, in-class administration of online instruments may be possible in computer labs or with student-owned laptops (Nulty, 2008). A computer lab or two in each building could serve as a ClassEval hub. Instructors could bring their students to the lab at a scheduled time, or it could be open for SET purposes all the time. If advertised, students would know where they could find an open computer and complete their SETs

during that time. Students without internet access or without the opportunity to spend time outside of class completing SETs would benefit from this administration method, potentially reducing nonresponse due to opportunity costs.

Advertisement of SETs should saturate the university environment - online and physically. Including the website address (URL) in emails, on the course web page, and perhaps on computers in computer labs (tangible signs and reminders on the computers' desktops) would continuously reinforce and encourage response (Nulty, 2008). These informal reminders could also increase students' opportunities to complete SETs. When they normally would have forgotten, reminders would reinforce SET submission.

Incentives have garnered support in survey administration, because paper surveys tend to increase response rates by offering money, gift cards, and other gifts, no matter how small. While some researchers discovered that web surveys with incentives do not produce results very different from those without incentives (Sax et al., 2003), others found that incentives based on non-monetary rewards are helpful to increase response rates (Anderson et al., 2005; Dommeyer et al., 2004; McGourty et al., 2002a, 2002b). For example, a one-quarter increase in students' grades if two-thirds of the class completed the SET for the course increased response rates in one study (Dommeyer et al., 2004). Several universities now require students to complete their SETs before releasing their semester grades (McGourty et al., 2002a, 2002b; Thorpe, 2002).

Whatever the incentive may be, it may influence response rates and perceptions of other campus surveys. Though incentives may be popular for paper surveys (Dillman et al., 2009; Groves et al., 2004, 2009) and increasingly explored for SETs and other

postsecondary surveys of students (Nulty, 2008; Porter, 2004), there may be negative consequences. Porter (2004) suggests that if incentives and other alternative avenues of raising response rates (e.g. multiple reminders, better communication with faculty, etc.) produce equal results, the alternative is the better choice. Incentives are not only expensive, but they may also become an expectation, decreasing response rates for other campus surveys.

Additional Institutional Practices

An increasing trend in survey administration offices on college and university campuses is the regulation of any surveys conducted on campus. Without this coordination and oversight, students receive multiple surveys about the same topics, and poorly designed surveys have the potential to permeate student inboxes. Moreover, many surveys of students attempt to collect information that another office or department already has. Data collection is a difficult process and in many cases could be avoided if institutional researchers or survey offices are aware of survey requests. This oversight needs further research, but may be effective in reducing survey fatigue.

Another practice for institutions to explore is the practice of formative SETs (Seldin, 1984, 1999, 2006). Because students are typically unable to see results and changes from summative SETs, knowing how the university utilizes SET results tends to increase response rates (Dommeyer et al., 2004; Zúñiga, 2004). Moreover, instructors can tune-in to their students' needs and respond accordingly, a characteristic of good instructional habits and teaching practices. Formative SETs should increase salience as well; if students see changes based on their response to formative SETs, end-of-course

evaluations may also be more salient and see increased response rates. They know their opinions matter and that feedback works.

Assumptions that students have access to online SETs may need a second or third glance. Technology is not always dependable, they might not always have access to it, and they may not feel they are able to give their opinions. Efforts to examine technology use and availability to students may uncover that they need further support. Previously mentioned methods to address these issues include computer lab SET administration and orientating students to SETs and completing them with useful feedback and confidence. In addition, demonstration of how to complete an SET (Dommeyer et al., 2004) and where computers with internet connections are available would also be helpful to students lacking adequate technology or knowledge about it.

Finally, as discussed in the beginning of this chapter, institutions need to consider the error that this study uncovered. Nonresponse error, especially nonresponse bias and the potential for double bias, tends to reduce data quality. However, patterns of nonresponse existed, and institutions may be able to target them through campus organizations and academic disciplines. Lab courses may also introduce over-coverage, because students evaluate the course and the lab in two separate instruments. Survey fatigue associated with multiple SETs for the same instructor or course may discourage response.

Future Research

This institution may be unlike others, and similar nonresponse studies on different campuses would be helpful to better understand the population of nonrespondents and

factors contributing to likelihood of response. Few studies have addressed SET nonresponse, and most of those that exist may be obsolete. Not only do nonrespondents' characteristics continue to vary, but also our use of technology continues to evolve. Moreover, nonresponse studies should not be a one-time occurrence; instead, replications would continue to track the evolving population of nonrespondents.

Limitations and Advantages

This study was able to utilize several variables underrepresented in previously conducted research on postsecondary campus surveys. The outcomes uncovered here deserve further study on this and other campuses. Academic environment, for the course and for the student, influenced response. However, the kinds of differences were inconsistent with previous studies (Porter & Whitcomb, 2005; Sax et al., 2003; Yu et al., 2007). These differences may be attributable to the institution, type of survey, or instructors, but they still contradict other studies. Therefore, I highly recommend replications of this study on this and other campuses to identify unresponsive populations and respond accordingly.

Additional limitations to this study included a restricted setting (NC State only), the exclusion of graduate students, and the exclusion of many desired characteristics of the course and instructor. The first two limitations are common and understandable. The restricted setting increases the applicability of the results to the institution, and excluding graduate students increases the generalizeability to other institutions as not all colleges and universities have these populations. However, including variables associated with instructors and other course-level predictors may provide a better understanding of

campus nonresponse. Extensions of this study should include these variables for a broader view of nonresponse. Other considerations for future research include item nonresponse, contacting nonrespondents for more information, and research validation of the effectiveness of institutional practices to combat nonresponse.

To extend the context, NC State is part of the University of North Carolina (UNC) system. By conducting similar studies at each of the 16 institutions associated with the UNC system, results could help institutional researchers better understand SET participation for the state and inform policy. Methods to combat nonresponse could be statewide, and policy could address the handling of results when low response rates abound.

Replication of this study would also be helpful to compare populations between semesters and validate results. Identifying nonresponse bias in one semester for one institution is not enough. After replication and validation, scheduled analyses of SET nonresponse should continue on this and all campuses. It is likely that nonrespondents' characteristics will change, and unless annual or biannual analyses can support previous findings, campuses may misdirect their efforts to combat nonresponse. It is important to remember that SETs are not just another survey; results influence personnel and curricular decisions made by administrators for the state of North Carolina. While survey research is applicable to the SET context, not all survey results can be as powerful.

Conclusions: No Evaluation Left Behind

Survey nonresponse is increasing (Dillman et al., 2009; Groves et al., 2004). As institutions increasingly change their methods of data collection from paper SETs to

online SETs, they are also seeing a decrease in response rates (Avery et al., 2006; Sax et al., 2003). To appropriately address and combat issues associated with nonresponse, this study made an important first step and identified the nonresponsive populations and factors that can contribute to the likelihood of each SET submission online.

At NC State for the Fall 2009 semester, respondents to SETs were typically female, with a GPA at or above 3.0, were older, were first-year or senior students, not African American or Asian, and were part of a realistic major. However, characteristics of the course also tended to influence participation. Submissions of SETs were more likely in the course and student's major were in the same department; an A, B, or C grade also increased the likelihood of response.

Salience and academic environment were the most helpful predictors, because they incorporated grades, majors, and course location. Applying nonresponse theories to this study increased the explanation and interpretation of outcomes. However, more information would further explain these outcomes. For example, survey fatigue, while significant, is holistically immeasurable. Survey researchers cannot measure or be aware of all surveys administered to individuals (e.g. restaurant, student services, consumer, and online shopping surveys).

To combat nonresponse, theories of survey participation can offer hope. SET administrators can combat survey fatigue and opportunity costs by altering the survey design and providing students with time and locations to complete SETs. Encouraging and training students to provide quality feedback uses social exchange to create respondents out of nonrespondents. Providing more information about the process creates

a culture of response. Specifically addressing nonresponse bias may take place within the student's academic environment, where response rates did not decrease for five of the six major types. Increasing response rates tends to increase data quality, and when surveys have high stakes, addressing nonresponse becomes a priority. Because postsecondary institutions continue to switch from paper to online SETs, contributions to research in the field of online SET participation will continue to grow and inform survey practices for colleges and universities.

REFERENCES

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*(2), 153.

Algozzine, B., Beattie, J., Bray, M., Flowers, C., Gretes, J., Howley, L., et al. (2004). Student evaluation of college teaching. *College Teaching 52*(4), 134–141.

Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Pine Forge Press.

American Council on Education. (1920). *The educational record*. Washington, D.C.: American Council on Education.

Anderson, H. M., Cain, J., & Bird, E. (2005). Online student course evaluations: Review of literature and a pilot study. *American Journal of Pharmaceutical Education, 69*(Article 5).

Astin, A. W., & Lee, C. (1967). Current practices in the evaluation and training of college teachers. In C. B. T. Lee (Ed.), *Improving college teaching* (pp. 361-375). Washington, D. C.: American Council on Education.

Astin, A. W. (1993). An empirical typology of college students. *Journal of College Student Development*, 34, 36–46.

Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic SETs: Does an online delivery system influence student evaluations? *Journal of Economic Education*, 21-37.

Baker, P. C., & Remmers, H. H. (1952). The measurement of teacher characteristics and the prediction of teaching efficiency in college. *Review of Educational Research, 22*(3), Teachers and Nonacademic Personnel, 169-174.

Ballantyne, C. (2003). Online evaluations of teaching: An examination of current practice and considerations for the future. In D. L. Sorenson & T. D. Johnson (Eds.), *Online student ratings of instruction: Vol. 96. New Directions for Teaching and Learning* (pp. 103-112). Hoboken, NJ: John Wiley & Sons.

Barr, A. S., Bechdolt, B. V., Gage, N. L., Orleans, J. S., Pace, C. R., Remmers, H. H., et al. (1953). Second report of the committee on criteria of teacher effectiveness. *The Journal of Educational Research, 46*(9), 641-658.

Basow, S. A. (1995). Student evaluation of college professors: When gender matters. *Journal of Educational Psychology, 87*(4), 656-665.

Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 71-86). New York: John Wiley & Sons.

Berinsky, A. J. (2008). Survey non-response. In W. Donsbach & M. W. Traugott (Eds.), *The SAGE handbook of public opinion research* (pp. 309-321). Thousand Oaks, CA: Sage Publications.

Berk, R. A. (2004). *Regression analysis: A constructive critique*. Thousand Oaks, CA: Sage Publications.

Blackburn, R. T., & Clark, M. J. (1975). An assessment of faculty performance: Some correlates between administrator, colleague, student and self-ratings. *Sociology of Education, 48*(2), 242-256.

Blau, P. (1964). *Exchange and power in social life*. New York: Wiley.

Boersma, F., Eefsting, J. A., van den Brink, W., & van Tilburg, W. (1997). Characteristics of non-respondents and the impact of non-response on prevalence estimates of dementia. *International Journal of Epidemiology, 26*(5), 1055-1062.

Bolman, L. G., & Deal, T. E. (2008). *Reframing organizations: Artistry, choice, and leadership* (4th ed.). San Francisco, CA: Jossey-Bass.

Brandenburg, G. C., & Remmers, H. H. (1927) The Purdue Rating Scale for Instructors. *Educational Administration and Supervision*, 13, 399-406.

Bresler, J. B. (1968). Teaching effectiveness and government awards. *Science, 160*(3824), 164-167.

Cashin, W. E., & Perrin, B. M. (1978). *IDEA technical report no.4: Description of IDEA Standard Form data base*. Manhattan: Kansas State University Center for Faculty Evaluation and Development.

Cashin, W. E. (1994). Student ratings of teaching: A summary of the research. In K. A. Feldman & M. B. Paulsen (Eds.), *Teaching and learning in the college classroom*, ASHE Reader Series. Needham Heights, MA: Ginn Press.

Centra, J. A. (1977). Student ratings of instruction and their relationship to student learning. *American Educational Research Journal, 14*(1), 17-24.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*(5), 495.

Charters, W. W. (1940). Sizing up the faculty. *The Journal of Higher Education, 11*(9), 457-461.

Clarksberg, M., Robertson, D., & Einarson, M. (2008). Engagement and student surveys: Nonresponse and implications for reporting survey data. Paper presented at the 48th Annual Forum of the Association for Institutional Research: Seattle, WA.

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn?: A meta-analysis and review of the literature. *Journal of Marketing Education, 31*(1), 16-30.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51*(3), 281-309.

Cohen, P. A. (1990). Bringing research into practice. In M. Theall & J. Franklin (Eds.), *Student ratings of instruction: Issues for improving practice: Vol. 43. New Directions for Institutional Research* (pp. 123-132). San Francisco, CA: Jossey-Bass.

Costin, F., Greenough, W. T., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research, 41*(5), 511-513.

Costin, F. (1978). Do student ratings of college teachers predict student achievement? *Teaching of Psychology, 5*(2), 86-88.

Couper, M. P., Kapteyn, A., Schonlau, M., & Winter, J. (2007). Noncoverage and nonresponse in an internet survey. *Social Science Research, 36*(1), 131-148.

Cramer, K. M., & Alexitch, L. R. (2000). Student evaluations of college professors: Identifying sources of bias. *Canadian Journal of Higher Education, 30*(2), 143-164.

Crumbley, D. L., & Fliedner, E. (2002). Accounting administrators perceptions of student evaluation of teaching (SET) information. *Quality Assurance in Education, 10*(4), 213-222.

Crumbley, D. L., & Reichelt, K. J. (2009). Teaching effectiveness, impression management, and dysfunctional behavior: Student evaluation of teaching control data. *Quality Assurance in Education, 17*(4), 377-392.

d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*(11), 1198-1208.

Davis, M., Hirschberg, J., Lye, J. N., & Johnston, C. G. (2007). Systematic influences on teaching evaluations: The case for caution. *Australian Economic Papers 46*(1), 18-38.

de Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 41-54). New York: John Wiley & Sons.

de Leeuw, J., & Meijer, E. (2008). *Handbook of multilevel analysis.* New York: Springer.

Detchen, L. (1940). Shall the student rate the professor? *The Journal of Higher Education, 11*(3), 146-154.

Dey, E. L. (1997). Working with low survey response rates: The efficacy of weighting adjustments. *Research in Higher Education, 38*(2), 215-227.

Dillman, D. A. (1978). *Mail and telephone surveys: The total design method*. New York, NY: John Wiley & Sons.

Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little, R. J. A. (2002). Survey nonresponse in design, data collection, and analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 3-26). New York: John Wiley & Sons.

Dillman, D.A., Smyth, J.D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. San Francisco: Jossey-Bass.

Dommeyer, C. J., Baum, P., Chapman, K. S. & Hanna, R. W. (2002a) Attitudes of business faculty toward two methods of collecting teaching evaluations: paper vs. online. *Assessment & Evaluation in Higher Education, 27*(5), 445–462.

Dommeyer, C. J., Baum, P., & Hanna, R. W. (2002b). College students' attitudes toward methods of collecting teaching evaluations: In-class versus on-line. *Journal of Education for Business, 78*(1), 11-15.

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment and Evaluation in Higher Education, 29*(5), 611-625.

Drucker, A. J., & Remmers, H. H. (1951). Do alumni and students differ in their attitudes toward instructors? *Journal of Educational Psychology*, 42, 129-143.

Duda, M. D., & Nobile, J. L. (2010). The fallacy of online surveys: No data are better than bad data. *Human Dimensions of Wildlife, 15*(1), 55-64.

Emerson, R. M. (1976). Social exchange theory. *Annual Review of Sociology,* 2, 335-362.

Etzioni, A. (1975). *A comparative analysis of complex organizations: On power, involvement, and their correlates*. New York: The Free Press.

Feldman, K. A. (1976a). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education, 4*(1), 69-111.

Feldman, K. A. (1976b). The superior college teacher from the students' view. *Research in Higher Education, 5*(3), 243-288.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education, 18*(1), 3-124.

Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30*(6), 583-645.

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-143). Dordrecht, Netherlands: Springer.

Fidelman, C. G. (2007). Course evaluation surveys: In-class paper surveys versus voluntary online surveys. Doctoral dissertation, UMI no. 3301790. Boston College.

Fite, D. (2006). Using evaluation data for personnel decisions. In P. Seldin (Ed.), *Evaluating faculty performance: A practical guide to assessing teaching, research, and service* (pp. 181-200). Bolton, MA: Anker Publishing Company, Inc.

Frick, T. W., Chadha, R., Watson, C., Wang, Y., & Green, P. (2009). College student perceptions of teaching and learning quality. *Educational Technology, Research and Development, 57*(5), 705-720.

Garson, G. D. (2009, December 16). *Linear mixed models*. Retrieved from http://faculty.chass.NC State.edu/garson/PA765/multilevel.htm.

Gage, N. L. (1974). Students' ratings of college teaching: Their justification and proper use. In N. S. Glasman & B. R. Killait (Eds.), *University of California at Santa Barbara conference on effective teaching*. Santa Barbara, CA: University of California.

Gamliel, E., & Davidovitz, L. (2005). Online versus traditional teaching evaluation: Mode can matter. *Assessment & Evaluation in Higher Education, 30*(6), 581-592.

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. (1978). The generalizeability of student ratings of instruction: Estimation of the teacher and course components. *Journal of Educational Measurement, 15*(1), 1-13.

Goyder, J., Warriner, K., & Miller, S. (2002). Evaluating socio-economic status (SES) bias in survey nonresponse. *Journal of Official Statistics, 18*(1), 1–11.

The Graduate Record Examinations. (1977, July). *How universities evaluate faculty performance: A survey of department heads* (GRE Board Research Report GREB No.75-5bR). Berkeley, CA: John A. Centra.

Greenwald, A.G. (1995). Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching. Paper presented at the annual meeting of the American Psychological Association, New York City.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley & Sons.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70*(5), 646-675.

Groves, R. M., Cialdini, R. B., & Couper, M. P (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56*,* 475-495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley & Sons.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation. *Public Opinion Quarterly*, 64, 299-308.

Groves, R. M., Dillman, D.A., Eltinge, J. L., & Little, R. J. A. (2002). *Survey nonresponse.* New York: John Wiley & Sons.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology.* Hoboken, NJ: John Wiley & Sons.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2-31.

Groves, R. M. Couper, M., Presser, S., Singer, E., Tourangeau, R., Piani Acosta, G., & Nelson, L. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly, 70*(5), 720-736.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly, 72*(2), 167-189.

Gustad, J. W. (1961). Policies and practices in faculty evaluation. *Educational Record,* 42, 194-211.

Gustad, J. W. (1967). Evaluation of teaching performance: Issues and possibilities. In C. B. T. Lee (Ed.), *Improving college teaching* (pp. 265-282). Washington, D.C.: American Council on Education.

Guthrie, E. R. (1953). The evaluation of teaching. *The American Journal of Nursing, 53*(2), 220-221.

Guthrie, E. R. (1954). *The evaluation of teaching: A progress report*. Seattle: University of Washington.

Haskell, R. E. (1997). Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st century. *Education Policy Analysis Archives, 5*(6). Retrieved from: http://epaa.asu.edu/ojs/article/view/607.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review, 43*(4), 447-462.

Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.

Heilman, J. D., & Armentrout, W. D. (1936). The rating of college teachers on ten traits by their students. *Journal of Educational Psychology, 27*(3), 197-216.

Hmieleski, K. & Champagne, M. V. (2000, September/October). Plugging in to course evaluation. *The Technology Source*. Retrieved from: The archives at the University of North Carolina http://technologysource.org/article/plugging_in_to_course_evaluation/.

Ho, D. E., & Shapiro, T. H. (2008). Evaluating SETs: An empirical analysis of a quasi-experiment at the Stanford Law School. *Journal of Legal Education 58*(3).

Hofstede, G. (1981). Culture and organizations. *International Studies of Management and Organization, 10*(4), 15-41.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, 6, 35-45.

Holland, J. L. (1966). *The psychology of vocational choice*. Waltham, MA: Blaisdell.

Holland, J. L (1973). *Making vocational choice: A theory of careers.* Englewood Cliffs, NJ: Prentice Hall.

Holland, J. L. (1985). *Making vocational choice: A theory of vocational personalities and work environments.* Englewood Cliffs, NJ: Prentice Hall.

Holland, J. L. (1997). *Making vocational choice: A theory of vocational personalities and work environments.* Odessa, FL: Psychological Assessment Resources.

Hox. J., & de Leeuw, E. (2002). The influence of interviewer's attitudes and behavior on household survey nonresponse: An international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 103-120). New York: John Wiley & Sons.

Isely, P., & Singh, H. (2005). Do higher grades lead to favorable student evaluations? *Journal of Economic Education, 36*(1), 29-42.

Johnson, T. D. (2003). Online student ratings: Will students respond? In D. L. Sorenson & T. D. Johnson (Eds.), *Online student ratings of instruction: Vol. 96. New Directions for Teaching and Learning* (pp. 49-59). Hoboken, NJ: John Wiley & Sons.

Johnson, T. P., O'Rourke, D., Burris, J., & Owens, L. (2002). Culture and survey nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 55-70). New York: John Wiley & Sons.

Jones, C. R. (2009). Nonresponse bias in online SETs. Doctoral dissertation, UMI no. 3386938. James Madison University.

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly, 68*(1), 94–101.

Kucsera, J. V., & Zimmaro, D. M. (2008). Electronic course instructor survey (eCIS) report. Austin, TX: Division of Instructional Innovation and Assessment, University of Texas at Austin.

Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. In M. Theall, Abrami, P. C., & Mets, L. A. (Eds.), *The student ratings debate: Are they valid? How can we best use them?: Vol. 109. New Directions for Institutional Research* (pp. 9-25). San Francisco, CA: Jossey-Bass.

Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), *Review of research in education, 3*(1), 210-240.

Langbein, L. (2005). Management by results: Student evaluation of faculty teaching and the mismeasurement of performance. Paper presented at the Public Choice Society Annual Meeting: New Orleans.

Laurie, H., Smith, R., & Scott, L. (1999). Strategies for reducing nonresponse in a longitudinal panel survey. *Journal of Official Statistics,* 15, 269-282.

Layne, B. H., DeCristoforo, J. R., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education, 40*(2), 221-232.

Lechuga, V. M. (2008). Assessment, knowledge, and customer service: Contextualizing faculty work at for-profit colleges and universities. *The Review of Higher Education*, *31*(3), 287-307.

Lepkowski, J. M., & Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 259-274). New York: John Wiley & Sons.

Leslie, L.L. (1972). Are high response rates essential to valid surveys? *Social Science Research 1*(3), 323-334.

Lévi-Strauss, C. (1967). *Structural anthropology*. Garden City, NY: Doubleday.

Lindner, J. R., Murphy, T. H., & Briers, G. H. (2001). Handling nonresponse in social science research. *Journal of Agricultural Education*, *42*(4), 43-53.

Linsky, A. S., & Straus, M. A. (1975). Student evaluations, research productivity, and eminence of college faculty. *The Journal of Higher Education, 46*(1), 89-102.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Marcus, B., & Schutz, A. (2005). Who are the people reluctant to participate in research? Personality correlates of four different types of nonresponse as inferred from self- and observer ratings. *Journal of Personality, 73*(4), 1-26.

Marlin, J. W. (1987). Student perceptions of end-of-SETs. *The Journal of Higher Education, 58*(6), 704-716.

Marsh, H. W. (1982). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology, 74*(2), 264-279.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*(5), 707-754.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research,* 11*,* 253-388.

Marsh, H. W. (2007). Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An*

evidence-based perspective (pp. 319-383). Dordrecht, Netherlands: Springer.

Marsh, H. W., & Overall, J. U. (1979). Long-term stability of students' evaluations: A note on Feldman's consistency and variability among college students in rating their teachers and courses. *Research in Higher Education, 10*(2), 139-147.

Marsh, H. W., Overall, J. U., & Kesler, S. P. (1979). Class size, students' evaluations, and instructional effectiveness. *American Educational Research Journal, 16*(1), 57-70.

Marsh, H. W., & Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology, 72*(4), 468-475.

Marsh, H. W., & Hocevar, D. (1991a). The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching and Teacher Education, 7*(1), 9-18.

Marsh, H. W., & Hocevar, D. (1991b). Students' evaluations of teaching effectiveness: The stability of mean ratings of the same teachers over a 13-year period. *Teaching and Teacher Education, 7*(4), 303-314.

Marsh, H. W., & Dunkin, M. J. (1992) Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 8, pp. 143-233). New York: Agathon Press.

Marsh, H. W., & Roche, L.A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist,* 52, 1187–1197.

McGhee, D. E., & Lowell, N. (2003). Psychometric properties of student ratings of instruction in online and on-campus courses. In D. L. Sorenson & T. D. Johnson (Eds.), *Online student ratings of instruction: Vol. 96. New Directions for Teaching and Learning* (pp. 39-48). Hoboken, NJ: John Wiley & Sons.

McGourty, J., Scoles, K., & Thorpe, S. (2002a, June). Web-based student evaluation of instruction: Promises and pitfalls. Paper presented at the 42nd Annual Forum of the Association for Institutional Research: Toronto, ON, Canada.

McGourty, J., Scoles, K., & Thorpe, S. (2002b, November). Web-based course evaluation: Comparing the experience at two universities. Paper presented at the 32nd ASEE/IEEE Frontiers in Education Conference: Boston, MA.

McKeachie, W. J. (1969) Student ratings of faculty. *American Association of University Professors Bulletin*, 55, 439-444.

McKeachie, W .J. (1979). Student ratings of faculty: A reprise. *Academe*, 65, 384–397.

McKeachie, W. J. (1994). *Teaching tips: Strategies, research, and theory for college and university teachers* (9th ed.). Lexington, MA: D.C. Heath/Houghton-Mifflin.

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*(11), 1218-1225.

Miklich, D. R. (1969). An experimental validation study of the Purdue Rating Scale for Instruction. *Educational and Psychological Measurement, 29*(4), 963-967.

Moore, D. L., & Tarnai, J. (2002). Evaluating nonresponse error in mail surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 197-212). New York: John Wiley & Sons.

Morton-Williams, J. (1993). *Interviewer approaches*. Aldershot, England: Dartmouth Publishing.

Nilson, L.B. (2003). *Teaching at its best: A research-based resource* (2nd Ed.). New York: Anker Publishing Co.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education, 33*(3), 301-314.

Pambookian, H. S. (1974). Initial level of student evaluation of instruction as a source of influence on instructor change after feedback. *Journal of Educational Psychology, 66*(1), 52-56.

Porter, S. R. (2004a). Pros and cons of paper and electronic surveys. In S. R. Porter (Ed.), *Overcoming survey research problems: Vol. 121. New Directions for Institutional Research* (pp. 91-98). San Francisco, CA: Jossey-Bass.

Porter, S. R. (2004b). Raising response rates: What works? In S. R. Porter (Ed.), *Overcoming survey research problems: Vol. 121. New Directions for Institutional Research* (pp. 5-22). San Francisco, CA: Jossey-Bass.

Porter, S. R. (2005). Survey research policies: An emerging issue for higher education. In P. D. Umbach (Ed.), S*urvey research: Emerging issues: Vol. 127. New Directions for Institutional Research* (pp. 5-16). San Francisco, CA: Jossey-Bass.

Porter, S. R. (2009). Do college student surveys have any validity? Paper presented at the 2009 annual forum of the Association for the Study of Higher Education: Vancouver, BC, Canada.

Porter, S. R., & Whitcomb, M. E. (2004). Understanding the effect of prizes on response

rates. In S. R. Porter (Ed.), *Overcoming survey research problems: Vol. 121. New Directions for Institutional Research* (pp. 51-62). San Francisco, CA: Jossey-Bass.

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. In S. R. Porter (Ed.), *Overcoming survey research problems: Vol. 121. New Directions for Institutional Research* (pp. 63-74). San Francisco, CA: Jossey-Bass.

Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement, and personality. *Research in Higher Education, 46*(2), 127-152.

Porter, S. R. & Whitcomb, M. E. (2007). Mixed-mode contacts in web surveys: Paper is not necessarily better. *Public Opinion Quarterly, 71*(4), 635-648.

Porter, S. R., & Umbach, P. D. (2006a). Student survey response rates across institutions: Why do they vary? *Research in Higher Education, 47*(2). 229-247.

Porter, S. R., & Umbach, P. D. (2006b). College major choice: An analysis of personal-environment fit. *Research in Higher Education, 47*(4), 429-449.

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? *Quality Assurance in Education, 15*(2), 178-191.

Quinn, D. (2003). Valuing evaluation: A case study of professional development to support academic engagement in online evaluation processes and outcomes. Paper presented at the Evaluations and Assessment Conference: University of South Australia, Adelaide.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & du Toit, M. (2004). *HLM6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Remedios, R., & Lieberman, D. A. (2008). I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal, 34*(1), 91-115.

Remmers, H. H. (1928). The relationship between students' marks and students' attitudes toward instructors. *School and Society*, 28, 759-760.

Remmers, H. H. (1930). To what extent do grades influence student ratings of instructors? *Journal of Educational Research*, 21, 314-316.

Remmers, H. H. (1934). Reliability and halo-effect of high school and college students' judgments of their teachers. *Journal of Applied Psychology*, 18, 619-630.

Remmers, H. H., & Brandenburg, G. C. (1927). Experimental data on the Purdue Rating Scale for Instructors. *Educational Administration and Supervision*, 13, 519-527.

Remmers, H. H., & Gage, N. L. (1949). Reshaping educational policy. *Review of Educational Research, 19*(1, The Social Framework of Education), 77-89.

Renaud, R. D., & Murray, H. G. (1996). Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education,* 223-240.

Rogelberg, S. G., & Luong, A. (1998). Nonresponse to mailed surveys: A review and guide. *Current Directions in Psychological Science, 7*(2), 60-65.

Roose, H., Lievens, J., & Waege, H. (2007). The joint effect of topic interest and follow-up procedures on the response in a mail questionnaire: An empirical test of the leverage-saliency theory in audience research. *Sociological Methods & Research, 35*(3), 410-428.

Root, A. R. (1931). Student ratings of teachers. *The Journal of Higher Education, 2*(6), 311-315.

Rosen, D., Holmberg, K., & Holland, J. L. (1997). *The educational opportunities finder*. Odessa, FL: Psychological Assessment Resources.

Rotem, A., & Glasman, N. S. (1979). On the effectiveness of students' evaluative feedback to university instructors. *Review of Educational Research, 49*(3), 497-511.

Rubin, D. B., & Zanutto, E. (2002). Using matched substitutes to adjust for nonignorable nonresponse through multiple imputations. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 389-402). New York: John Wiley & Sons.

Rumberger, R. W., & Palardy, G. J. (2004). Multilevel models for school effectiveness research. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 235-258). Thousand Oaks, CA: Sage Publications.

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education, 44*(4), 409-432.

Sax, L. J., Gilmartin, S. K., Lee, J. J., & Hagedorn, L. S. (2008). Using web surveys to reach community college students: An analysis of response rates and response bias. *Community College Journal of Research & Practice, 32*(9), 712-729.

Scates, D. E., Baker, P. C., & Remmers, H. H. (1951). Progress in research on personnel evaluation. *Journal of Teacher Education, 2*(2), 143-146.

Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. *Research in Higher Education, 38*(5), 575-592.

Schwarz, N. (2008). The psychology of survey response. In W. Donsbach & M. W. Traugott (Eds.), *The SAGE handbook of public opinion research (pp. 374-387)*. Thousand Oaks, CA: Sage Publications.

Scott, S. V. (1999). The academic as service provider: Is the customer "always right"? *Journal of Higher Education Policy and Management*, *21*(2), 193-202.

Seldin, P. (1975). *How colleges evaluate professors: Current policies and practices in evaluating classroom teaching performance in liberal arts colleges.* New York: Blythe-Pennington, Ltd.

Seldin, P. (1984). *Changing practices in faculty evaluation*. San Francisco, CA: Jossey-Bass.

Seldin, P. (1999). *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions.* Bolton, MA: Anker Publishing Company, Inc.

Seldin, P. (2006). *Evaluating faculty performance: A practical guide to assessing teaching, research, and service.* Bolton, MA: Anker Publishing Company, Inc.

Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy: Markets, state, and higher education*. Baltimore, MD: The Johns Hopkins University Press.

Smart, J. C., Feldman, K. A., & Ethington, C. A. (1999). Major-field and personality-environment fit: Using Holland's theory to study change and stability of students. *The Journal of Higher Education, 70*(6), 642-669.

Smart, J. C., Feldman, K. A., & Ethington, C. A. (2000). *Academic disciplines: Holland's theory and the study of college students and faculty*. Nashville, TN: Vanderbilt University Press.

Stake, J. E. (1997). Response to Haskell: Academic freedom, tenure, and student evaluation of faculty. *Education Policy Analysis Archives, 5*(8). Retrieved from: http://epaa.asu.edu/ojs/article/view/609.

Stalnaker, J. M., & Remmers, H. H. (1928). Can students discriminate traits associated with success in teaching? *Journal of Applied Psychology, 12*(6), 602-610.

Stevens, J. J., & Aleamoni, L. M. (1984). The use of evaluative feedback for instructional improvement: A longitudinal perspective. *Instructional Science, 13*(4), 285-304.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York: Pearson Education, Inc.

Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education, 28*(3), 169-173.

Theall, M. (1997). On drawing reasonable conclusions about student ratings of instruction: A reply to Haskell and to Stake. *Education Policy Analysis Archives, 5*(Commentary).

Thorpe, S. W. (2002). Online student evaluation of instruction: An investigation of non-response bias. Paper presented at the 42nd Annual Forum for the Association for Institutional Research, Toronto, ON, Canada.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Uekawa, K. (2005). HLM. Retrieved from http://www.estat.us/id38.html.

Umbach, P. D. (2004). Web surveys: Best practices. In S. R. Porter (Ed.), *Overcoming survey research problems: Vol. 121. New Directions for Institutional Research* (pp. 23-38). San Francisco, CA: Jossey-Bass.

Umbach, P. D. (2005). Getting back to the basics of survey research. In P. D. Umbach (Ed.), S*urvey research: Emerging issues: Vol. 127. New Directions for Institutional Research* (pp. 91-100). San Francisco, CA: Jossey-Bass.

Umbach, P. D., & Milem, J. F. (2004). Applying Holland's typology to the study of differences in student views about diversity. *Research in Higher Education, 45*(6), 625-649.

Umbach, P. D., & Porter, S. R. (2002). How do academic departments impact student satisfaction? Understanding the contextual effects of departments. *Research in Higher Education, 43*(2, AIR Forum Issue), 209-234.

UPA. (2008a). Background. Retrieved June 23, 2009, from North Carolina State University's University Planning & Analysis Web site: http://www2.acs.NC State.edu/UPA/classeval/faculty/background.htm.

UPA. (2008b). Summary statistics: NC State University online class evaluation summary, Spring 2007-Spring 2008. Retrieved June 23, 2009, from North Carolina State University's University Planning & Analysis Web site: http://www2.acs.NC State.edu/UPA/classeval/faculty/Summary%20Statistics.pdf.

UPA. (2008c). Evaluating ClassEval. Retrieved June 23, 2009, from North Carolina State University's University Planning & Analysis Web site: http://www2.acs.NC State.edu/UPA/classeval/faculty/Combined_3.pdf.

UPA. (2008d). How the results are used. Retrieved June 30, 2009, from North Carolina State University's University Planning & Analysis Web site: http://www2.acs.NC State.edu/UPA/classeval/students/results.htm.

Walker, G. E., Golde, C. M., Jones, L., Bueschel, A. C., & Hutchings, P. (2008). *The formation of scholars: Rethinking doctoral education for twenty-first century*. San Francisco, CA: Jossey-Bass.

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: A practical guide to using statistical software*. New York: Chapman & Hall/CRC.

Williams, W. M., & Ceci, S. J. (1997). " How'm I doing?": Problems with student ratings of instructors and courses. *Change, 29*(5), 12-23.

Wilson, W.R. (1932). Students rating teachers. *The Journal of Higher Education, 3*(2), 75-82.

Vehovar, V., Batagelj, Z., Manfreda, K.L., & Zaletel, M. (2002). Nonresponse in web surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 229-242). New York: John Wiley & Sons.

Yu, C. H., Jannasch-Pennell, A., DiGangi, S., Kim, C., & Andrews, S. (2007). Data visualization and mining for survey responses. *Practical Assessment, Research & Evaluation, 12(19)*, 1-12.

Zanutto, E., & Zaslavsky, A. (2002). Using administrative records to impute for nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 403-416). New York: John Wiley & Sons.

Zúñiga, R. E. (2004, March). Increasing response rates for online surveys – a report from the Flashlight program's BeTA project. *F-LIGHT: E-newsletter for the flashlight program*. Retrieved from http://www.tltgroup.org/resources/F-LIGHT/2004/03-04.html#BeTA.

APPENDICES

Appendix A: ClassEval Instrument

NC STATE UNIVERSITY

Online Class Evaluation (ClassEval)

CSC 295 W 001 ST-APPLIED WEB PRO SMITH MARK HENRY
Instructor: SMITH MARK HENRY

The purpose of this confidential evaluation is to improve the quality of courses and teaching effectiveness by providing faculty with constructive student feedback. Please indicate your response to each item in the following sections by marking "Strongly Agree" to "Strongly Disagree" or "Not Applicable".

Questions related to Instructor and Course	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree	Not Applicable
1. The instructor stated course objectives/outcomes	O	O	O	O	O	O
2. The instructor was receptive to students outside the classroom	O	O	O	O	O	O
3. The instructor explained difficult material well	O	O	O	O	O	O
4. The instructor was enthusiastic about teaching the course	O	O	O	O	O	O
5. The instructor was prepared for class	O	O	O	O	O	O
6. The instructor gave prompt and useful feedback	O	O	O	O	O	O
7. The instructor effectively used instructional technology	O	O	O	O	O	O
Questions related to Instructor and Course	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree	Not Applicable
8. The instructor consistently treated students with respect	O	O	O	O	O	O
9. Overall, the instructor was an effective teacher	O	O	O	O	O	O
10. The course readings were valuable aids to learning	O	O	O	O	O	O
11. The course assignments were valuable aids to learning	O	O	O	O	O	O
12. This course was intellectually challenging and stimulating	O	O	O	O	O	O
13. This course improved my knowledge of the subject	O	O	O	O	O	O
14. Overall, this course was excellent	O	O	O	O	O	O

Answer Questions 15-18 ONLY if this evaluation is for a course that is exclusively lab based OR if it is for the lab section/instructor for a course that includes a separate lab potion.

Questions related to Labs/Lab Courses	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree	Not Applicable
15. Lab sessions contributed to mastery of course concepts	O	O	O	O	O	O
16. Lab facilities, equipment, supplies, etc. were adequate	O	O	O	O	O	O
17. The degree of lab difficulty was appropriate	O	O	O	O	O	O
18. Overall, the labs were effective learning experiences	O	O	O	O	O	O

Please answer the following open-ended questions in the text areas below.

19. Comment on strengths and weaknesses of the instructor.

20. Comment on strengths and weaknesses of the course.

21. Comment on strengths and weaknesses of the lab (if applicable)

22. Other comments

Appendix B: Original Request for Course Evaluation Completion

Students:

ClassEval – NC State's online system for evaluating your classes – is now open. Please go to https://classeval.ncsu.edu/ , login with your unity ID, and rate your courses and instructors. If you need help, write to classeval@ncsu.edu .

Your opinions really do help to make a difference. For example, Professor Parsons in Chemical and Biomolecular Engineering says, "In previous incarnations of [one] class, I used homework and class presentations as a means to determine student grades. Several students indicated on the evaluations that class exams would help them to better comprehend the class materials. I now use frequent short tests in the class, and my evaluation feedback indicates that students react positively to these."

Please be sure to complete all of your evaluations. The University takes this important activity very seriously - - so between now and deadline for ClassEval we will continue to send you email reminders until your complete the evaluation for each of your courses.

Rest assured that your responses are confidential. The class average rating and comments will not be given to the instructor until after grades are posted. However, be aware that if your class is very small (less than 5 students) there is a greater chance that your instructor will be able to identify your responses. In larger classes, your instructor will not know how any individual student responded.

ClassEval is open through 8 am on Wednesday, December 9. For more information, visit http://www.ncsu.edu/UPA/classeval/.

Appendix C: Fall 2009 Email Reminder Requests

Time Sent	Total	% of All	Cumlative %
Thu., Nov. 19	9,126	6.12	6.12
Fri., Nov. 20	2,708	1.82	7.94
Sat., Nov. 21	3,339	2.24	10.18
Sun., Nov. 22	3,120	2.09	12.27
Mon., Nov. 23	6,267	4.20	16.47
Tue., Nov. 24	2,021	1.35	17.82
Wed., Nov. 25	709	0.48	18.30
Thu., Nov. 26	430	0.29	18.59
Fri., Nov. 27	2,399	1.61	20.20
Sat., Nov. 28	1,291	0.87	21.07
Sun., Nov. 29	1,816	1.22	22.29
Mon., Nov. 30	7,944	5.32	27.61
Tue., Dec. 1	4,610	3.09	30.70
Wed., Dec. 2	5,536	3.71	34.41
Thu., Dec. 3	5,160	3.46	37.87
Fri., Dec. 4	2,090	1.40	39.27
Sat., Dec. 5	3,153	2.11	41.38
Sun., Dec. 6	2,210	1.48	42.86
Mon., Dec. 7	5,872	3.94	46.80
Tue., Dec. 8	7,428	4.98	51.78
Wed., Dec. 9	796	0.53	52.31
Total	78,025	52.30	

Note: Original email was sent on Thursday, November 19. Highlighting indicates all invitations and reminders sent to students, requesting completion of course evaluations.

Figure C1: **Daily response totals, percentages, and reminders.**

Time Sent	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Total	% of All	Cumlative %
Thu., Nov. 19	0	0	0	0	0	0	0	0	223	1,156	1,523	946	989	769	620	565	456	334	376	292	292	231	204	150	9,126	6.12	
Fri., Nov. 20	123	46	34	8	6	15	10	38	83	175	284	281	257	248	212	154	165	86	104	103	77	87	73	39	2,708	1.82	7.94
Sat., Nov. 21	17	19	11	6	0	1	2	1	15	33	33	55	71	103	375	579	487	316	232	246	237	196	167	137	3,339	2.24	10.18
Sun., Nov. 22	96	27	14	13	5	5	7	12	22	52	121	220	196	211	221	247	214	230	213	157	242	228	191	176	3,120	2.09	12.27
Mon., Nov. 23	99	37	27	13	12	7	18	22	326	623	730	627	655	506	377	343	379	324	228	189	256	194	133	142	6,267	4.20	16.47
Tue., Nov. 24	86	26	16	4	6	5	7	25	62	204	271	173	201	146	131	115	128	89	39	49	46	83	72	37	2,021	1.35	17.82
Wed., Nov. 25	10	1	12	7	0	0	5	11	15	44	21	46	69	78	26	56	39	19	35	51	45	45	28	46	709	0.48	18.30
Thu., Nov. 26	17	13	4	0	2	0	0	0	5	10	24	16	53	32	24	24	19	15	19	17	40	35	32	29	430	0.29	18.59
Fri., Nov. 27	11	9	1	7	1	0	2	6	6	14	36	39	215	415	283	269	179	174	115	134	115	120	119	129	2,399	1.61	20.20
Sat., Nov. 28	51	40	17	13	0	4	4	14	39	48	87	53	83	81	103	96	82	81	67	70	69	73	63	53	1,291	0.87	21.07
Sun., Nov. 29	55	38	16	0	3	4	1	4	23	20	56	87	105	142	101	97	112	96	101	123	206	158	179	89	1,816	1.22	22.29
Mon., Nov. 30	52	43	11	1	1	3	6	48	299	599	786	744	677	703	650	608	485	407	385	361	318	317	220	220	7,944	5.32	27.61
Tue., Dec. 1	92	90	22	17	9	2	16	22	191	162	351	238	437	470	239	419	356	212	254	175	237	257	194	148	4,610	3.09	30.70
Wed., Dec. 2	97	47	32	22	3	0	13	48	77	173	318	814	581	468	368	433	387	261	240	277	227	241	240	169	5,536	3.71	34.41
Thu., Dec. 3	124	51	10	5	3	4	13	21	66	297	661	608	551	457	420	384	261	252	177	191	140	196	149	119	5,160	3.46	37.87
Fri., Dec. 4	52	22	12	8	13	5	3	32	78	156	268	210	165	182	144	131	183	114	54	44	37	64	54	59	2,090	1.40	39.27
Sat., Dec. 5	31	12	5	0	1	0	6	38	85	147	214	321	340	326	277	227	193	197	180	125	143	118	83	84	3,153	2.11	41.38
Sun., Dec. 6	64	41	44	16	4	3	1	10	14	45	88	90	141	179	193	152	161	138	109	141	148	171	120	137	2,210	1.48	42.86
Mon., Dec. 7	83	37	31	8	2	1	8	48	265	340	554	613	593	469	526	406	357	325	161	180	307	210	181	167	5,872	3.94	46.80
Tue., Dec. 8	124	79	38	27	19	5	11	29	207	376	573	577	592	650	487	481	407	408	343	377	383	389	428	418	7,428	4.98	51.78
Wed., Dec. 9	293	126	93	75	31	32	52	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	796	0.53	52.31
																								Total	78,025	52.30	

Figure C2: **Hourly response totals, percentages, and reminders.**

Appendix D: Text of Follow-up Reminders Sent Via Email

1) Students:
ClassEval – NC State's online system for evaluating your classes – is now open. Please go to https://classeval.ncsu.edu/ , login with your unity ID, and rate your courses and instructors. If you need help, write to classeval@ncsu.edu .

Your opinions really do help to make a difference. For example, Professor Parsons in Chemical and Biomolecular Engineering says, "In previous incarnations of [one] class, I used homework and class presentations as a means to determine student grades. Several students indicated on the evaluations that class exams would help them to better comprehend the class materials. I now use frequent short tests in the class, and my evaluation feedback indicates that students react positively to these."

Please be sure to complete all of your evaluations. The University takes this important activity very seriously - - so between now and deadline for ClassEval we will continue to send you email reminders until your complete the evaluation for each of your courses.

Rest assured that your responses are confidential. The class average rating and comments will not be given to the instructor until after grades are posted. However, be aware that if your class is very small (less than 5 students) there is a greater chance that your instructor will be able to identify your responses. In larger classes, your instructor will not know how any individual student responded.

ClassEval is open through 8 am on Wednesday, December 9. For more information, visit http://www.ncsu.edu/UPA/classeval/.

2) With just a few minutes of your time, you can help make NC State a better university. By evaluating your classes through ClassEval, you'll give instructors valuable feedback they can use to improve their teaching and course/s.

Professor Martha Crowley says: "I've used evaluations to determine where I need to expand the number of class periods devoted to a particular book, where I've needed to incorporate more class discussion and current events, and the kinds of changes I need to make in writing quizzes and exams. Just as importantly, I've learned from positive feedback where *not* to make changes. Evaluations tell me not just where I need to make adjustments, but also what I'm doing that's working for students.

We will stop sending you email reminders as soon as you complete all evaluations. Please go to https://classeval.ncsu.edu/, login with your unity ID, and rate all your instructors and courses. If you need help, write to classeval@ncsu.edu .

Your responses are entirely confidential. Your instructors will see only the average class scores and written comments without any names attached - AFTER they have submitted grades for the semester.

ClassEval is open until 8 am on Wednesday, December 9. For more information, visit http://www.ncsu.edu/UPA/classeval/ .

3) With just a few minutes of your time, you can help make NC State a better university. By evaluating your classes through ClassEval, you'll give instructors valuable feedback they can use to improve their teaching and course/s.

Please go to https://classeval.ncsu.edu , login with your unity ID, and rate all your instructors and courses. After completing the evaluations for all your classes we will stop sending you email reminders. If you need help, write to classeval@ncsu.edu .

Your responses are entirely confidential. Your instructors will see only the average class scores and written comments without any names attached - AFTER they have submitted grades for the semester.

ClassEval closes 8 am on Wednesday, December 9 when finals begin. For more information, visit http://www.ncsu.edu/UPA/classeval/ .

4) Subject: Make this your last ClassEval message

By this time next week, you'll be thinking of finals. So don't wait to evaluate all your instructors and classes! ClassEval closes 8 am on Wednesday, December 9 just before finals begin.

URL: https://classeval.ncsu.edu
Login: Your Unity ID
Help: Write to classeval@ncsu.edu

If you've taken statistics, you know that a high response rate is necessary to make your class evaluations into meaningful feedback. Your responses are 100% confidential. For more information, visit http://www2.acs.ncsu.edu/UPA/classeval/index.htm.

Your name will be excluded from ClassEval email reminders once all your classes and instructors evaluations have been submitted.

Thanks for your participation.

5) Subject: Who will win?

Students:

Will graduate students lead the undergrads again this semester? Will students in the First Year College continue to outdo other freshmen? How about women and international students -- will they still outperform the men and US citizens? Will CHASS and Management students respond more than Textiles and Design?

It's about ClassEval -- and who finishes the evaluations that faculty use to improve their courses. C'mon! Make a difference at State!

Go to: https://classeval.ncsu.edu/
Find help at classeval@ncsu.edu

Monitor your college's response rate at
http://classeval.ncsu.edu/applications/dashboard_coll.cfm?reportstyle=3 and your classes' response rate at
http://classeval.ncsu.edu/applications/dashboard.cfm?reportstyle=3

You will stop receiving email reminders once all your surveys are complete. ClassEval ends Wednesday, December 9 at 8 am just before finals begin.

Thanks!
ClassEval Staff

6) Something different, just for you!

http://www.youtube.com/watch?v=Q0Ci3f-_b0o

Go to https://classeval.ncsu.edu/
Find help at classeval@ncsu.edu

Thanks!
ClassEval Staff

7) Subject: Who is in the lead?

Students:

Will graduate students lead the undergrads again this semester? Will students in the First Year College continue to outdo other freshmen? How about women and international students -- will they still outperform the men and US citizens? Will CHASS and Management students respond more than Textiles and Design?

First Year College students are still 'way ahead. Design students still lag the crowd.
Students in the social sciences and humanities are doing better than engineers and scientists.

It's about ClassEval -- and who finishes the evaluations that faculty use to improve their courses. C'mon! Make a difference at State!

Go to https://classeval.ncsu.edu/
Find help at classeval@ncsu.edu
Monitor your college's response rate at
http://classeval.ncsu.edu/applications/dashboard_coll.cfm?reportstyle=3 and your classes' response rate at
http://classeval.ncsu.edu/applications/dashboard.cfm?reportstyle=3

Thanks!
ClassEval Staff

8) ..Reasons to Evaluate Your Classes

10. The tables will be turned...you will be grading your teacher!
9. Teachers and their bosses really do read your comments. You can tell them what you really think...it's anonymous!
8. Studies show that an instructor's happiness is correlated with high participation rates in ClassEval.
7. You want to do the right thing and help make the course better for students next semester.
6. You can help your favorite teacher get a raise or teaching award.
5. You can give advice to your professor, like: get a new textbook, or put your syllabus online.
4. The Interim Provost sent you an email asking you to do it (although with his cool Australian accent, it might have been better if he'd phoned).
3. You can take a breather from studying for finals for a few minutes. It's less fattening than another trip to the fridge.
2. You will stop receiving annoying ClassEval reminders.
1. If you don't speak up, who will?

https://classeval.ncsu.edu

9) Classes are over! There are less than 24 hours to finish ClassEval. This is the last call for evaluations.

Your instructors regularly evaluate your work. It's your turn to evaluate them at https://classeval.ncsu.edu/ . Now is the time before finals start.

Please take a few minutes to share your opinions about your instructors. Your advice plays a key role in rewarding good teaching and in identifying needed improvements. Your ratings and comments are entirely confidential.

Please finish all evaluations by 8 am on Wednesday, December 9. Contact mailto:classeval@ncsu.edu if you need help. Thank you for your participation and helping NC State become a better university.

10) Classes are over! There are less than 24 hours to finish ClassEval. This is the last call for evaluations.

Your instructors regularly evaluate your work. It's your turn to evaluate them at https://classeval.ncsu.edu/ . Now is the time before finals start.

Please take a few minutes to share your opinions about your instructors. Your advice plays a key role in rewarding good teaching and in identifying needed improvements. Your ratings and comments are entirely confidential.

Please finish all evaluations by 8 am on Wednesday, December 9. Contact mailto:classeval@ncsu.edu if you need help. Thank you for your participation and helping NC State become a better university.

Appendix E: Email Sent to Faculty About ClassEval

To all faculty teaching in Fall 2009:

ClassEval, the University's online class evaluation system, will be open from 8 am on November 19 until 8 am on December 9 when finals begin. During this period, students will receive email messages asking them to evaluate each of their classes and labs (with a few exceptions noted below) by going to a website and completing brief questionnaires. When the evaluation period begins, please direct your students to the ClassEval URL: https://classeval.ncsu.edu/

If they need help, they should contact classeval@ncsu.edu . You can write this information on the board and/or distribute this information on paper. We mailed your department posters to hang in classrooms and on hallway bulletin boards. Please encourage your students as strongly as you can to complete a questionnaire for your course. Our research on ClassEval suggests that undergraduates and students enrolled through distance education may need more prompting than graduate students and on-campus courses, respectively. If your classroom has good wireless connectivity you may want to encourage students to bring laptops and provide some class time to complete the survey. The Evaluation of Teaching Committee and the Faculty Senate strongly discourage giving students any kind of incentives for completing the evaluations. Incentives can skew the data and compromise confidentiality. Instead, give examples of how you have used the information to improve your teaching, and explain how the results are used by your department to make promotion and tenure decisions.

Students are more inclined to participate if they believe their effort might make a difference. Lists of classes included in and excluded from this process are posted on the ClassEval website at http://www2.acs.ncsu.edu/UPA/classeval/courses_list.htm . Once ClassEval opens, you can monitor the response rate for each of your classes by watching a dashboard posted at http://classeval.ncsu.edu/applications/dashboard.cfm?reportstyle=3.

As soon as possible after the ClassEval period is over and all grades have been posted (after December 20), confidential reports will be posted on a protected website accessible only by you and your department head. Instructors may find reports at https://classeval.ncsu.edu/instrrep/courses2.cfm , and heads may find reports at https://classeval.ncsu.edu/dynarep/index.cfm . All data - both instructors' ratings and students' responses - are confidential. Individually identifiable information about teaching evaluations will be reported only to the instructor and department head as a confidential, personnel matter. Instructors will not know how individual students responded. All data, including individual students' responses and the resulting data are confidential and protected by the same federal and state laws, and by the same university policies, as student and personnel information, which means that they may be reported publicly only in the aggregate. Reminders are sent every few days to students who have not completed all the class evaluations. Students will stop receiving the reminders as soon as they have completed all their class evaluations. Questions and comments may be sent to lewis_carson@ncsu.edu .

Appendix F: Additional ClassEval Reminders (Not Emails)

1. Placed an ad in the Technician (Student Newspaper)
2. MyPack portal popup reminder
3. Signs across campus
4. Youtube Video: http://www.youtube.com/watch?v=Q0Ci3f-_b0o

Appendix G: Institutional Review Board (IRB) Application

North Carolina State University
Institutional Review Board for the Use of Human Subjects in Research
SUBMISSION FOR NEW STUDIES

GENERAL INFORMATION

1. **Date Submitted:** *4 March 2010*
1a. **Revised Date:** _____
2. **Title of Project:** *Nonresponse to ClassEval*
3. **Principal Investigator:** *Meredith J.D. Adams*
4. **Department:** *Leadership, Policy, and Adult & Higher Education*
5. **Campus Box Number:** *7801*
6. **Email:** *mjadams2@ncsu.edu*
7. **Phone Number:** *919.513.7928*
8. **Fax Number:** *n/a*
9. **Faculty Sponsor Name and Email Address if Student Submission:** *Paul Umbach , paul_umbach@ncsu.edu*
10. **Source of Funding? (required information):** *none*
11. **Is this research receiving federal funding?:** *no*
12. **If Externally funded, include sponsor name and university account number:** *n/a*
13. **RANK:**
    - ☐ Faculty
    - ☒ Student: ☐Undergraduate; ☐ Masters; or ☒PhD
    - ☐ Other (specify): _____

As the principal investigator, my signature testifies that I have read and understood the University Policy and Procedures for the Use of Human Subjects in Research. I assure the Committee that all procedures performed under this project will be conducted exactly as outlined in the Proposal Narrative and that any modification to this protocol will be submitted to the Committee in the form of an amendment for its approval prior to implementation.

Principal Investigator:

Meredith J.D. Adams	*Meredith J.D. Adams* *	*4 March 2010*
(typed/printed name)	(signature)	(date)

*As the faculty sponsor, my signature testifies that I have reviewed this application thoroughly and will oversee the research in its entirety. I hereby acknowledge my role as the **principal investigator of record**.*

Faculty Sponsor:

Paul Umbach	Copied on Submission Email *	_____
(typed/printed name)	(signature)	(date)

***Electronic submissions to the IRB are considered signed via an electronic signature. For student submissions this means that the faculty sponsor has reviewed the proposal prior to it being submitted and is copied on the submission.**

Please complete this application and email as an attachment to: debra_paxton@ncsu.edu or send by mail to: Institutional Review Board, Box 7514, NCSU Campus (Administrative Services III). **Please include consent forms and other study documents with your application and submit as one document.**

****

For SPARCS office use only

Reviewer Decision (Expedited or Exempt Review)

☐ Exempt ☐ Approved ☐ Approved pending modifications ☐ Table

Expedited Review Category: ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8a ☐ 8b ☐ 8c ☐ 9

Reviewer Name	Signature	Date

North Carolina State University
Institutional Review Board for the Use of Human Subjects in Research
GUIDELINES FOR A PROPOSAL NARRATIVE

In your narrative, address each of the topics outlined below. Every application for IRB review must contain a proposal narrative, and failure to follow these directions will result in delays in reviewing/processing the protocol.

A. INTRODUCTION

1. Briefly describe in lay language the purpose of the proposed research and why it is important.

The purpose of this proposed research is to understand the population of students at North Carolina State University that does not respond to their end of course evaluations and examine the potential bias introduced by nonresponse.

This is important for several reasons. First, nonresponse error, a critical error that survey researchers face when conducting any survey, occurs when respondents' characteristics differ from those who did not respond. About 10,000 undergraduate students do not submit their course evaluations, increasing the risk of nonresponse error and nonresponse bias. Second, survey participation is on a steady decline, and surveys administered to postsecondary students are no exception. This also increases the risk of bias and error associated with nonresponse. These first two issues can led to outcomes represented in the third reason this study is important. In the age of accountability and data-driven decision-making processes, nonresponse can severely affect data quality and lead to incorrect conclusions. NCSU utilizes results from end-of-course evaluations as components of personnel records and for promotion and tenure decisions, and instructors can utilize them for instructional improvement. Fourth, if this population of nonresponders is better understood, the university can utilize the proper techniques to decrease nonresponse rates and therefore the quality of results.

2. If student research, indicate whether for a course, thesis, dissertation, or independent research.

This study is for a presentation at the annual forum for the Association for Institutional Research (May 2010), as well as topic of a dissertation study.

B. SUBJECT POPULATION

1. How many subjects will be involved in the research?

The population will be the undergraduates from the fall semester who were eligible for completing online evaluations of their course(s). UPA (NCSU's institutional research office) has the data, and current numbers are unknown since data will not be released until we obtain IRB approval.

2. Describe how subjects will be recruited. Please provide the IRB with any recruitment materials that will be used.

Students are requested to fill out online course evaluations at the end of every semester/term. Undergraduate students who take courses that are subject to course evaluations will be included in the dataset.

3. List specific eligibility requirements for subjects (or describe screening procedures), including those criteria that would exclude otherwise acceptable subjects.

Undergraduate students who take courses that are eligible according to UPA guidelines course evaluations will be included in the dataset (http://www2.acs.ncsu.edu/UPA/classeval/class_exclusions.htm). To protect students' identity in smaller courses, there are no evaluations for classes with less than six students.

4. Explain any sampling procedure that might exclude specific populations.

Students in courses with less than six enrolled do not complete course evaluations; so that data is not included. Therefore, students in majors or programs known for very small class sizes might not be represented, unless they are taking other classes that are larger. Courses and students that are not considered to be at the undergraduate level, such as graduate students, graduate courses, post-baccalaureate studies students, and non-credit courses, will not be included. This addresses the reliability of the study, producing results more generalizeable and consistent. (http://www2.acs.ncsu.edu/UPA/classeval/class_exclusions.htm)

5. Disclose any relationship between researcher and subjects - such as, teacher/student; employer/employee.

The only relationship that exists is the one I have with student data in the College of Education. I am a research assistant in the college IR office "Knowledge Management and Assessment" and work regularly with sensitive data from students, faculty, and staff. I am under a strict confidentiality agreement to keep all information with which I work safe, secure, and undisclosed. While I am involved with Master's and Doctoral level courses with my assistantship sponsor, I do not teach undergraduate courses, and undergraduates are the population that is the focus of this study.

6. Check any vulnerable populations included in study:

- ☒ minors (under age 18) - if so, have you included a line on the consent form for the parent/guardian signature
- ☐ fetuses
- ☐ pregnant women
- ☐ persons with mental, psychiatric or emotional disabilities
- ☐ persons with physical disabilities
- ☒ economically or educationally disadvantaged
- ☐ prisoners
- ☒ elderly
- ☐ students from a class taught by principal investigator
- ☒ other vulnerable population. - potentially

7. If any of the above are used, state the necessity for doing so. Please indicate the approximate age range of the minors to be involved.

Economically disadvantaged may be found in the variable "If the student receives financial aid", though that category may or may not fit many students. Research demonstrates that financial status may bias nonresponse (Berinksy, 2008; Groves, Singer, & Corning, 2000; Porter & Whitcomb, 2005).

Vulnerability may be possible for any demographic or characteristic depending on the individual and the situation. In this case, if a certain population is discovered not to participate in a university tradition, they may feel vulnerable to the other populations who do participate.

Elderly and minor persons may be included since one variable to be included is age. However, at this time it is unknown to the researcher if any undergraduate students exist in these categories. Some studies report that the elderly are less likely to provide answers in many surveys (Boersma, Eefsting, Van Der Brink, & Tilburg, 2002; Dillman, Etinge, Groves, & Little, 2002).

It is important to know that the rest of these characteristics of vulnerable populations remain unknown to the researcher. Though some students may have some of these characteristics, the researcher will not know. (For example, pregnant women are most likely included in the population, but this information is not included in the dataset.) Identifiable data will be excluded from the dataset to be provided by the UPA office. Neither personal interviews nor any contact with undergraduate students are included in the methodology of this study, and no names, Unity IDs, emails, or other such directly identifiable information will be in the dataset.

PROCEDURES TO BE FOLLOWED

1. In lay language, describe completely all procedures to be followed during the course of the experimentation. Provide sufficient detail so that the Committee is able to assess potential risks to human subjects. In order for the IRB to completely understand the experience of the subjects in your project, please provide a detailed outline of everything subjects will experience as a result of participating in your project. Please be specific and include information on all aspects of the research, through subject recruitment and ending when the subject's role in the project is complete. All descriptions should include the informed consent process, interactions between the subjects and the researcher, and any tasks, tests, etc. that involve subjects. If the project involves more than one group of subjects (e.g. teachers and students, employees and supervisors), please make sure to provide descriptions for each subject group.
2.

1. At the end of each semester, students receive emails from the UPA office to complete their course

evaluations. Although unity IDs are required to complete these evaluations, UPA keeps student information and identifies completely anonymous from any outside source. Reminder emails (and sometimes reminders from professors and others) are periodically sent to discourage nonresponse.

2. After two weeks, the course evaluation period ends, and the UPA office prepares reports from course evaluation results for the faculty, keeping student identities confidential.
3. For this study, UPA will prepare a dataset that includes characteristic variables of students who were eligible to take their end-of-course evaluations (see list of variables below). This dataset will only include variables that will not identify any student. After several meetings with this office, it was discussed that the variables below were both unidentifiable and obtainable; this list was sent to the researcher from the UPA office.
4. Upon IRB approval, UPA will be able to complete the dataset.
5. The data collection already completed by UPA does not involve experimentation or rigorous data collection for the researcher. It only utilizes a de-identified dataset provided by the university's institutional research office (UPA). Since April 2009, the researcher and a faculty member have been working with the UPA office to build an institutional study that only utilizes variables that keep students' identities hidden. The UPA statistician ran data through several tests to ensure that the data I would receive pending IRB approval is not directly revealing of any identities. (Variables are listed below). No real data collection will take place, with the exception of obtaining the dataset from the UPA office pending IRB approval.
6. After receipt of the dataset, efforts to become familiar with it. Descriptive statistics will be reported in aggregate form.
7. Using Hierarchical Linear Modeling, student characteristics will be utilized to discover any bias they may have as a nonrespondent population.
8. Results will be in aggregate form, and single cases will never be included in the report.

The following variables will be in the dataset:

1. Semester – semester of course
2. Student Key – unique but non-traceable student identification number
3. Course Key – unique but non-traceable course identification code
4. Response Rate – percent of 18 core survey questions answered
5. Semester Response Rate – percent of surveys student filled out at least one question
6. Age – student age at semester Census (10th day of classes)
7. Gender – gender of student
8. Race - race
9. Semester GPA – combined GPA for all courses taken that semester
10. Course GPA – GPA for the individual course
11. Transfer Student – did student attend another post-secondary institution before NC State
12. Student Level – level of undergraduate student by hours passed
13. Total SC H – total credit hours passed at all institutions that NC State accepts
14. NCSU SCH – total credit hours passed at NC State
15. Semester SCH - credit hours taken in semester
16. Nationality – student's nationality
17. Major – student's primary major on Census (10th day of classes)
18. Minor – student's primary minor on Census
19. Entry Semester – first semester that student entered NC State at undergraduate level
20. Semesters Enrolled – number of semester attended course at NC State (excludes summer school)
21. Historic Response Rate – number of surveys where student completed at lease one question in past
22. Financial Aid – did student receive financial aid in this semester
23. Course Same as Student – is course's college same as student's primary major college
24. Number of Courses – number of courses taken that semester
25. Athlete – is student a scholarship or non-scholarship student-athlete
26. Campus Housing – is student in campus housing
27. TOFL – did student take TOFL exam
28. SAT – highest SAT score combination
29. Residency – NC resident for tuition purposes
30. Response Time - time survey completed
31. Response Date – date survey completed

3. How much time will be required of each subject?

None- Course evaluations are not required by NCSU, and I will not conduct any follow-up on the nonresponders.

D. POTENTIAL RISKS

1. State the potential risks (physical, psychological, financial, social, legal or other) connected with the proposed procedures and explain the steps taken to minimize these risks.

1. Students may be identifiable – the UPA office is working to create a dataset excluding variables that could contribute to the identification of students while containing variables that keep students unidentifiable. A high number of cases in the dataset also contributes to student confidentiality. Also, the results and statistical modeling by nature do not report individual statistics. Any descriptives or results will not display identifiable characteristics. No attempts to make connections will be made.
2. Faculty may be identifiable – for this reason, no faculty or course-specific information is included in the dataset.
3. Personnel records may be breached – this study does not seek to obtain any faculty information. It also does not want to see the ratings and results from the course evaluations - only student characteristics will be in the dataset.

2. Will there be a request for information that subjects might consider to be personal or sensitive (e.g. private behavior, economic status, sexual issues, religious beliefs, or other matters that if made public might impair their self-esteem or reputation or could reasonably place the subjects at risk of criminal or civil liability)?

Several variables might be considered personal or sensitive. The following variables are scheduled to be in the dataset:

1. Semester – semester of course
2. Student Key – unique but non-traceable student identification number
3. Course Key – unique but non-traceable course identification code
4. Response Rate – percent of 18 core survey questions answered
5. Semester Response Rate – percent of surveys student filled out at least one question
6. Age – student age at semester Census (10th day of classes)
7. Gender – gender of student
8. Race - race
9. Semester GPA – combined GPA for all courses taken that semester
10. Course GPA – GPA for the individual course
11. Transfer Student – did student attend another post-secondary institution before NC State
12. Student Level – level of undergraduate student by hours passed
13. Total SC H – total credit hours passed at all institutions that NC State accepts
14. NCSU SCH – total credit hours passed at NC State
15. Semester SCH - credit hours taken in semester
16. Nationality – student's nationality
17. Major – student's primary major on Census (10th day of classes)
18. Minor – student's primary minor on Census
19. Entry Semester – first semester that student entered NC State at undergraduate level
20. Semesters Enrolled – number of semester attended course at NC State (excludes summer school)
21. Historic Response Rate – number of surveys where student completed at lease one question in past
22. Financial Aid – did student receive financial aid in this semester
23. Course Same as Student – is course's college same as student's primary major college
24. Number of Courses – number of courses taken that semester
25. Athlete – is student a scholarship or non-scholarship student-athlete
26. Campus Housing – is student in campus housing
27. TOFL – did student take TOFL exam
28. SAT – highest SAT score combination
29. Residency – NC resident for tuition purposes
30. Response Time - time survey completed
31. Response Date – date survey completed

a. If yes, please describe and explain the steps taken to minimize these risks.

Some variables may be considered personal or sensitive; however, identities are hidden from the researcher. The above variables, contained in a dataset that will be created by UPA (University Planning and Analysis), will be placed into statistical models to better understand the nonrespondent population of students. None of these characteristics includes name, Unity ID, student number, or other directly identifiable characteristic. Results will be reported in aggregate form. The researcher will not share the data with anyone outside of the research team/committee. Data will be kept in a safe, secure office in Poe Hall, which is locked at all hours.

Several variables originally included in the study have been excluded from consideration by UPA, because they may lend to the possible identification of students and/or faculty.

b. Could any of the study procedures produce stress or anxiety, or be considered offensive, threatening, or degrading? If yes, please describe why they are important and what arrangements have been made for handling an emotional reaction from the subject.

The potential risks listed above may produce these effects:

1. Students may feel pressure and stress to complete these course evaluations, because they already have a lot to do at the end of a semester. Perhaps this is why these evaluations are optional. These stress and anxiety possible are issues not associated with this study, but rather associated with the course evaluations they complete every semester/term.
2. Students may be identifiable – the UPA office is working to create a dataset excluding variables that could contribute to the identification of students while containing variables that keep students unidentifiable. A high number of cases in the dataset also contributes to student confidentiality. Also, the results and statistical modeling by nature do not report individual statistics. Any descriptives or results will not display identifiable characteristics. No individual-styled results or analysis will be conducted
3. The variables may demonstrate pronounced distinctions among groups. Populations of students discovered to respond more or less may feel pressure from other groups because of their choice. Groups may feel discrimination or feel that results may contribute to a stereotype. However, this is the potential risk for any study that includes such characteristics. It is important to note in any report that this information is not reflective of everyone in a particular population. The modeling will show the probability that a certain population will respond and explain who the typical nonresponder is.
4. Faculty may be identifiable – for this reason, no faculty or course-specific information is included in the dataset. (see variable list)
5. Personnel records may be breached – this study does not seek to obtain any faculty information. It also does not want to see the ratings and results from the course evaluations - only student characteristics will be in the dataset.

3. How will data be recorded and stored?

The dataset, provided by the university's UPA office, will be kept in a secure office location in Poe Hall (Room 520L) that has been designated a secure office location due to the nature of the data already housed there. It will be on a password-protected computer, which is on a secure server. Files will also be password protected.

a. How will identifiers be used in study notes and other materials?

Student identification will not be disclosed to this study, as such variables will not be provided by UPA. De-identified student characteristics will be in the dataset, but study notes and other materials will be useful only in aggregate form. No listing of characteristics for an individual case will be created, and such notes or materials would be unnecessary to this study.

b. How will reports will be written, in aggregate terms, or will individual responses be described?

Individual responses or characteristics will not be identified or described. Results will be reported in aggregrate terms and no single case or cases will be considered.

4. If audio or videotaping is done how will the tapes be stored and how/when will the tapes be destroyed at the conclusion of the study.

N/A

5. Is there any deception of the human subjects involved in this study? If yes, please describe why it is necessary and describe the debriefing procedures that have been arranged.

No.

E. POTENTIAL BENEFITS

This does not include any form of compensation for participation.

1. What, if any, direct benefit is to be gained by the subject? If no direct benefit is expected, but indirect benefit may be expected (knowledge may be gained that could help others), please explain.

There are three possible benefits:

1. Data quality potentially increases. If response rates increase, the likelihood for nonresponse error decreases (Dillman, et al., 2002; Groves, Dillman, Etinge, & Little, 2002; Groves et al., 2004).
2. Better data may result in better decisions. Decreasing the risk for error means increasing the risk for better quality data. This can increase administrators' abilities to make better-informed, higher-quality decisions that affect the institution, programs, departments, colleges, and faculty members in their promotion, tenure, and reappointment reviews. In addition, instructors can make better decisions about improving their teaching methods, and the potential to accomplish the purposes of the course evaluations increases.

   According to UPA (http://www2.acs.ncsu.edu/UPA/classeval/students/results.htm), there are five purposes of course evaluations at NCSU: 1) instructor use the results for improvement of teaching and courses, 2) department heads use the information to assign courses and… 3) …evaluate instructors for salary increases and promotions, 4) a summary is in faculty portfolios to be considered for tenure, 5) the information can be used to apply for grants or awards. (Also found in Feldman, 2007; Marsh; 2007).
3. Student satisfaction and quality assurance. In the literature, measurements of customer satisfaction and quality assurance at institutions of higher education are common (Costin et al., 1971; Feldman, 2007; Langbein, 2005; Slaughter & Rhoades, 2004; Wang, 1975). Course evaluations can be part of this measure. Through this study, the university may be able to understand their populations of nonresponders better. They then can employ methods to reach these populations so all students understand the importance of completing their course evaluations. In turn, students may increase their likelihood of participating in these surveys and having their voices heard in the process. Their institution can better serve them if they make their opinions and perspectives to faculty and administrators.

F. COMPENSATION

Please keep in mind that the logistics of providing compensation to your subjects (e.g., if your business office requires names of subjects who received compensation) may compromise anonymity or complicate confidentiality protections. If, while arranging for subject compensation, you must make changes to the anonymity or confidentiality provisions for your research, you must contact the IRB office prior to implementing those changes.

1. Describe compensation

none

2. Explain compensation provisions if the subject withdraws prior to completion of the study.

Not applicable

3. If class credit will be given, list the amount and alternative ways to earn the same amount of credit.

Not applicable

G COLLABORATORS

1. If you anticipate that additional investigators (other than those named on **Cover Page**) may be involved in this research, list them here indicating their institution, department and phone number.

1. **Members of the UPA office**
2. **My committee/dissertation team: (all are NCSU employees)**

Paul Umbach – Assoc Prof in Leadership, Policy, and Adult and Higher Education Dept: paul_umbach@ncsu.edu
Betsy Brown – Vice Provost for Faculty Affairs 919-513-7741
Bonnie Fusarelli – Assoc Prof in Leadership, Policy, and Adult and Higher Education Dept 919 515-6359
James Bartlett – Assoc Prof in Leadership, Policy, and Adult and Higher Education Dept 919.515.624
Gerald Ponder – Associate Dean of the College of Education 919.515.5906

2. Will anyone besides the PI or the research team have access to the data (including completed surveys) from the moment they are collected until they are destroyed.

No one else will be given access to the data. My office computer is password protected. I share an office with Malina Monaco (Director of Knowledge Management and Assessment in the College of Education). Poe Hall 520L has been assigned to this department as a safe and secure location, since we house institutional, college, personnel, and departmental data.

H. CONFLICT OF INTEREST

1. Do you have a significant financial interest or other conflict of interest in the sponsor of this project? *No*

2. Does your current conflicts of interest management plan include this relationship and is it being properly followed? *N/A*

I. ADDITIONAL INFORMATION

1. If a questionnaire, survey or interview instrument is to be used, attach a copy to this proposal. *University's survey instrument for ClassEval is below.*

2. Attach a copy of the informed consent form to this proposal. *Not Applicable*

3. Please provide any additional materials that may aid the IRB in making its decision.

I am fortunate to have the UPA office willing to help the formation of a research study, especially in the creation of a dataset with unidentifiable variables. Karen Helm and Trey Standish have been particularly helpful, meeting with me several times in working to form this project. Nancy Whelchel and Lewis Carson have also been part of the formation of the research plan.

J. HUMAN SUBJECT ETHICS TRAINING:

*Please consider taking the Collaborative Institutional Training Initiative (CITI), a free, comprehensive ethics training program for researchers conducting research with human subjects. Just click on the underlined link. *I have completed the human subject ethics training model and other ethics modules throughout graduate coursework and workshops.*

Appendix H: IRB Exempt Letter

North Carolina State University is a land-grant university and a constituent institution of The University of North Carolina

Office of Research and Graduate Studies

NC STATE UNIVERSITY

Sponsored Programs and
Regulatory Compliance
Campus Box 7514
2701 Sullivan Drive
Raleigh, NC 27695-7514

919.515.2444
919.515.7721 (fax)

From: Carol Mickelson, IRB Coordinator
North Carolina State University
Institutional Review Board

Date: March 12, 2010

Project Title:Nonresponse to ClassEval

IRB#: 1372-10

Dear Ms. Adams;

The research proposal named above has received administrative review and has been approved as exempt from the policy as outlined in the Code of Federal Regulations (Exemption: 46.101. b.2). Provided that the only participation of the subjects is as described in the proposal narrative, this project is exempt from further review.

NOTE:

1. This committee complies with requirements found in Title 45 part 46 of The Code of Federal Regulations. For NCSU projects, the Assurance Number is: FWA00003429.

2. Any changes to the research must be submitted and approved by the IRB prior to implementation.

3. If any unanticipated problems occur, they must be reported to the IRB office within 5 business days.

Please forward a copy of this letter to your faculty sponsor, if applicable. Thank you.

Sincerely,

Carol Mickelson
NCSU IRB

Appendix I: Majors Associated with Holland's Six Types

Table I1: **Major/department matched with personality/environment type.**

Major/Department	*Holland's Type**
Agriculture Education	Social
UNDECIDED: Animal sciences	Realistic
Animal Science General	Realistic
Agriculture and Resource Economics	Investigative
Agricultural Biological Engineering	Investigative
Biochemistry	Investigative
Biological Sciences	Investigative
Dean's Office - CALS	Investigative
Food science	Social
Horticulture science	Realistic
Microbiology	Investigative
Plant Biology	Investigative
Poultry Science	Realistic
Sociology - Criminology/Forensics	Investigative
Zoology	Investigative
Design General	Realistic
Architecture	Investigative
Graphic design	Realistic
Industrial Design	Artistic
Landscape Architecture	Artistic
Curriculum and Instruction	Enterprising
Education (Including Math/Science Ed)	Social
Biomedical Engineering	Realistic
Civil Engineering	Investigative
Chemical Engineering	Investigative
Computer Science	Investigative
Electrical and Computer Engineering	Realistic
Dean's Office - Engineering	Investigative
Industrial Engineering	Investigative
Mechanical and Aerospace Engineering	Realistic
Materials Engineering	Investigative
Nuclear Engineering	Investigative
Dean's Office - Natural Resources	Enterprising
Forestry General	Realistic

Table I1: **(continued).**

Major/Department	*Holland's Type**
Parks and Recreation	Enterprising
Wood and Paper Science	Realistic
Communications General	Enterprising
English General	Artistic
Foreign Language	Investigative
History General	Enterprising
UNDECIDED: CHASS	Social
Inter Disciplinary Studies	Social
Philosophy/Religion	Artistic
Political Science	Enterprising
Psych General	Investigative
Sociology/Anthropology	Investigative
Social Work	Social
Chemistry General	Investigative
Math General	Investigative
Physical Science (Other)	Investigative
Dean's Office - PAMS	Investigative
Physics	Investigative
Statistics (General)	Investigative
Textile Science and Engineering	Investigative
Textile and Apparel Management	Investigative
Textile Chemistry	Investigative
Accounting	Conventional
Business Admin Management General	Enterprising
Economics General	Investigative
Dean's Office - Management	Investigative
First Year College	Undecided
Transition Program	Undecided

* Majors categorized according to primary type in D. Rosen, K. Holmberg, & J. L. Holland's. *The Educational Opportunities Finder* (1997).

Appendix J: Additional Descriptive Statistics of Variables

Table J1: **Level-two variables.**

		Range			
	N	*Min.*	*Max.*	*Mean*	*SD*
Number of Evaluations Administered	134,929	1*	23*	5.96*	2.023*
	22,639*				
SAT Score	19,612	600	1600	1178.64	134.413
Number of semesters at NCSU	22,639	1	13	4.23	2.689
On Campus housing	22,639	0	1	.33	.470
Double major	22,639	0	1	.05	.213
Athlete	22,639	0	1	.03	.160
Undecided major	22,639	0	1	.12	.321
Female	22,639	0	1	.44	.496
Transfer	22,639	0	1	.16	.366
Resident	22,639	0	1	.92	.277
Part-time	22,639	0	1	.06	.230
GPA_in_2s	22,639	0	1	.42	.493
GPA_in_1s	22,639	0	1	.06	.238
GPAbelow1	22,639	0	1	.01	.095
SOPH	22,639	0	1	.23	.420
Junior	22,639	0	1	.24	.426
Senior	22,639	0	1	.28	.451
INTL student	22,639	0	1	.01	.099
Trad Student Age	22,639	0	1	.94	.243
African American	22,639	0	1	.09	.281
Other	22,639	0	1	.04	.191
Asian	22,639	0	1	.05	.227
Did Not Disclose RACE	22,639	0	1	.05	.212
Has SAT Score	22,639	0	1	.87	.340

* Number of or per students (not evaluations).

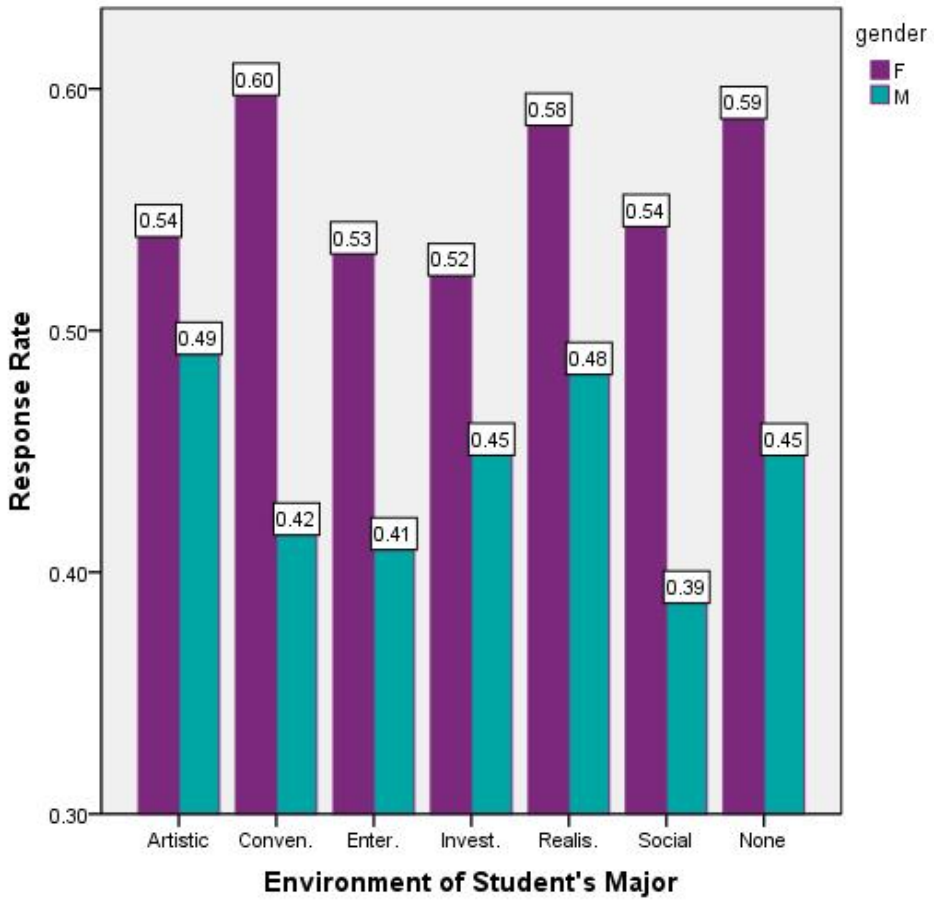


***Figure J1*: Response rate within environment: By gender.**

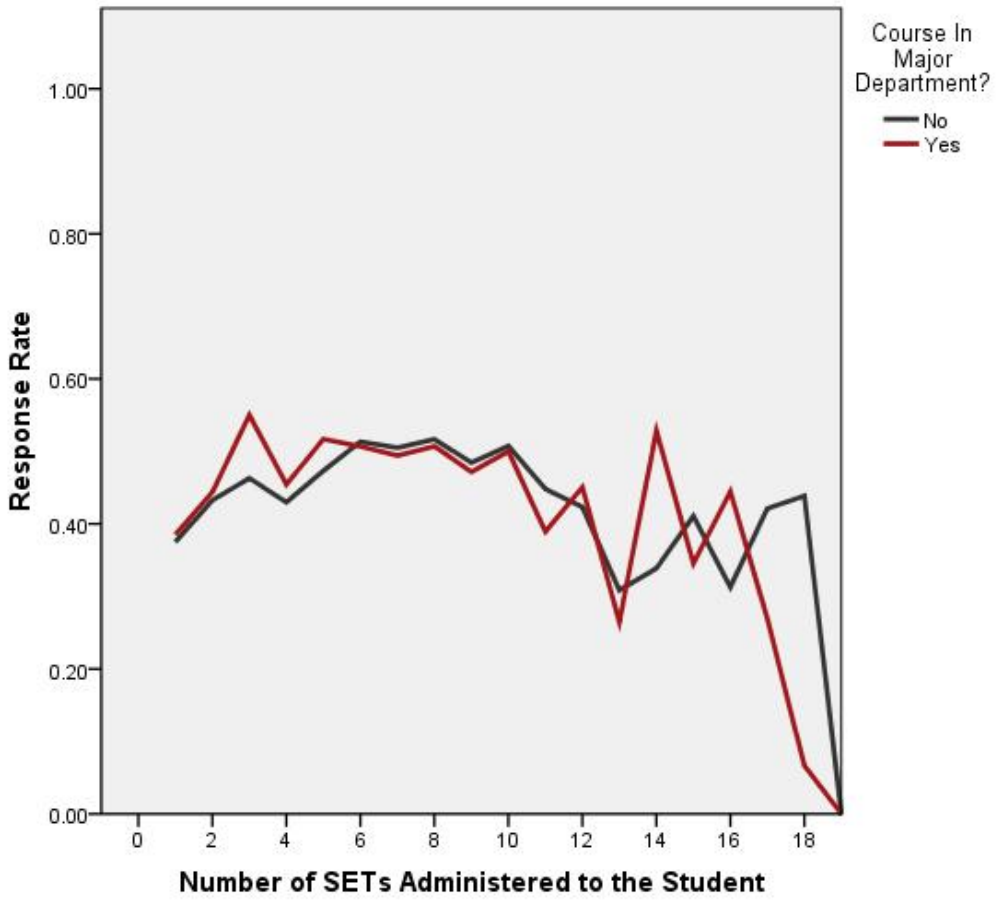


Figure J2: **Survey fatigue for course location.**

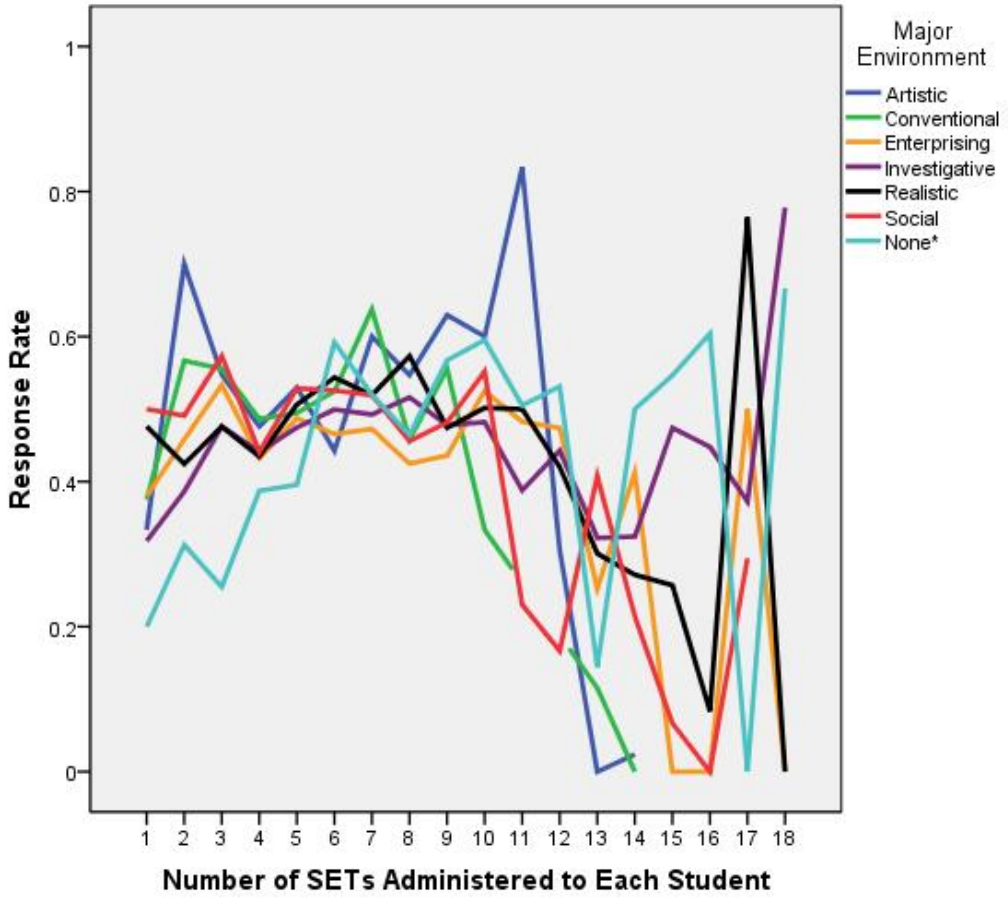


Figure J3: **Survey fatigue in academic environment.**

Appendix K: Fully Controlled Model, No Interactions

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$\text{Prob}(\text{URESPYN}=1\,\pi) = \varphi$$

$$\text{Log}[\varphi/(1 - \varphi)] = \eta$$

$$\eta = \pi_0$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\begin{aligned}\pi_0 = {} & \beta_{00} + \beta_{01}(\boldsymbol{NCSUCH}) + \beta_{02}(\boldsymbol{SEMCH}) + \beta_{03}(\boldsymbol{CAMPUSYN}) + \beta_{04}(\boldsymbol{ATHLYN}) + \beta_{05}(\boldsymbol{INTL}) + \\ & \beta_{06}(\boldsymbol{SOPH}) + \beta_{07}(\boldsymbol{JUNIOR}) + \beta_{08}(\boldsymbol{SENIOR}) + \beta_{09}(\boldsymbol{GPAIN2S}) + \beta_{010}(\boldsymbol{GPAIN1S}) + \\ & \beta_{011}(\boldsymbol{GPALESS1}) + \beta_{012}(\boldsymbol{TRNSFRYN}) + \beta_{013}(\boldsymbol{RESYN}) + \beta_{014}(\boldsymbol{AFRAMER}) + \\ & \beta_{015}(\boldsymbol{ASIAN}) + \beta_{016}(\boldsymbol{OTHRRACE}) + \beta_{017}(\boldsymbol{RACEUNRF}) + \beta_{018}(\boldsymbol{TRADAGE}) + \\ & \beta_{019}(\boldsymbol{PARTYN}) + \beta_{020}(\boldsymbol{FEMALEYN}) + \beta_{021}(\boldsymbol{ARTISTIC}) + \beta_{022}(\boldsymbol{CONVENT}) + \\ & \beta_{023}(\boldsymbol{ENTERPRI}) + \beta_{024}(\boldsymbol{INVESTIG}) + \beta_{025}(\boldsymbol{REALISTI}) + \beta_{026}(\boldsymbol{FYCTRANS}) + r_0\end{aligned}$$

Figure K1: **Fully controlled model equation.**

***Table K1*: Fully controlled model without interactions.**

Variable	*Controlled Model*	
	Coeff.	*Odds Ratio*
Intercept	-0.063***	0.939
NCSU CH	-0.002*	0.998
Semester CH	0.016*	1.016
On campus Housing	0.231***	1.260
Athlete	-0.382***	0.682
International Student	0.231	1.259
Class Rank *(First-year = ref.)*		
Sophomore	-0.247***	0.782
Junior	-0.183**	0.832
Senior	-0.013	0.987
Academic Environment *(Social = ref.)*		
Artistic	0.058	1.059
Conventional	0.087	1.091
Enterprising	-0.028	0.972
Investigative	0.000	1.000
Realistic	0.183***	1.200
FYC/Transitions	0.132*	1.142
Race *(White/Caucasian = ref.)*		
African American	-0.226***	0.798
Asian	-0.154**	0.857
Other Race	-0.111	0.895
Race Unknown	0.007	1.007
Semester GPA *(3.0-4.0 = ref.)*		
Semester GPA in 2s	-0.446***	0.640
Semester GPA in 1s	-1.198***	0.302
Semester GPA less than 1	-2.230**	0.108
Transfer Student	0.180***	1.197
NC Resident	0.019	1.019
Traditional Age (< 25 years)	-0.362***	0.696
Female	0.351***	1.421
Part-time	-0.041	0.960
Tau	6.059	
Intercept Reliability	0.735	

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

Appendix L: Hypothesis Two Models: Grade Earned

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$\text{Prob}(\text{URESPYN}=1 \mid \pi) = \varphi$$

$$\text{Log}[\varphi/(1 - \varphi)] = \eta$$

$$\eta = \pi_0 + \pi_1(\mathbf{GRADEB}) + \pi_2(\mathbf{GRADEC}) + \pi_3(\mathbf{GRADED}) + \pi_4(\mathbf{GRADESF}) + \pi_5(\mathbf{GRADEF}) + \pi_6(\mathbf{GRADENG})$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\pi_0 = \beta_{00} + \beta_{01}(\mathbf{NCSUCH}) + \beta_{02}(\mathbf{SEMCH}) + \beta_{03}(\mathbf{CAMPUSYN}) + \beta_{04}(\mathbf{ATHLYN}) + \beta_{05}(\mathbf{INTL}) + \beta_{06}(\mathbf{SOPH}) + \beta_{07}(\mathbf{JUNIOR}) + \beta_{08}(\mathbf{SENIOR}) + \beta_{09}(\mathbf{GPAIN2S}) + \beta_{010}(\mathbf{GPAIN1S}) + \beta_{011}(\mathbf{GPALESS1}) + \beta_{012}(\mathbf{TRNSFRYN}) + \beta_{013}(\mathbf{RESYN}) + \beta_{014}(\mathbf{AFRAMER}) + \beta_{015}(\mathbf{ASIAN}) + \beta_{016}(\mathbf{OTHRRACE}) + \beta_{017}(\mathbf{RACEUNRP}) + \beta_{018}(\mathbf{TRADAGE}) + \beta_{019}(\mathbf{PARTYN}) + \beta_{020}(\mathbf{FEMALEYN}) + \beta_{021}(\mathbf{ARTISTIC}) + \beta_{022}(\mathbf{CONVENT}) + \beta_{023}(\mathbf{ENTERPRI}) + \beta_{024}(\mathbf{INVESTIG}) + \beta_{025}(\mathbf{REALISTI}) + \beta_{026}(\mathbf{FYCTRANS}) + r_0$$

$$\pi_1 = \beta_{10} + r_1$$

$$\pi_2 = \beta_{20} + r_2$$

$$\pi_3 = \beta_{30} + r_3$$

$$\pi_4 = \beta_{40} + r_4$$

$$\pi_5 = \beta_{50} + r_5$$

$$\pi_6 = \beta_{60} + r_6$$

Figure L1: **Equation for grade-earned model.**

$$\eta = \beta_{00} + \beta_{01}*NCSUCH + \beta_{02}*SEMCH + \beta_{03}*CAMPUSYN + \beta_{04}*ATHLYN + \beta_{05}*INTL + \beta_{06}*SOPH + \beta_{07}*JUNIOR + \beta_{08}*SENIOR + \beta_{09}*GPAIN2S + \beta_{010}*GPAIN1S + \beta_{011}*GPALESS1 + \beta_{012}*TRNSFRYN + \beta_{013}*RESYN + \beta_{014}*AFRAMER + \beta_{015}*ASIAN + \beta_{016}*OTHRRACE + \beta_{017}*RACEUNRP + \beta_{018}*TRADAGE + \beta_{019}*PARTYN + \beta_{020}*FEMALEYN + \beta_{021}*ARTISTIC + \beta_{022}*CONVENT + \beta_{023}*ENTERPRI + \beta_{024}*INVESTIG + \beta_{025}*REALISTI + \beta_{026}*FYCTRANS + \beta_{10}*GRADEB + \beta_{20}*GRADEC + \beta_{30}*GRADED + \beta_{40}*GRADESF + \beta_{50}*GRADEF + \beta_{60}*GRADENG + r_0$$

Figure L2: **Mixed model equation: See Figure L1.**

Table L1: **Fully controlled model: Hypothesis two for all students.**

Variable	***Controlled Model***	
	Coeff.	*Odds Ratio*
Level One *(Grade A = ref. group)*		
Student Grade = B	-0.004	0.996
Student Grade = C	-0.038**	0.963
Student Grade = D	-0.064**	0.938
Student Grade = F/U	-0.422***	0.656
Student Grade = Satisfactory (S)	-0.351***	0.704
No Grade	-0.488***	0.614
Level Two	-0.071***	0.931
Artistic✢	0.054	1.056
Conventional✢	0.081	1.085
Enterprising✢	-0.039	0.962
Investigative✢	0.029	1.029
Realistic✢	0.207***	1.230
FYC/Transitions✢	0.123*	1.131
Sophomore✢✢	-0.239***	0.788
Junior✢✢	-0.196**	0.822
Senior✢✢	-0.032	0.968
NCSU CH	-0.002*	0.998
Semester CH	0.016*	1.016
Female	0.352***	1.422
On campus Housing	0.231***	1.260
Athlete	-0.392***	0.676
International Student	0.232	1.261
Transfer Student	0.184***	1.202
NC Resident	0.018	1.018
Traditional Age (< 25 years)	-0.380***	0.684
Part-time	-0.043	0.958
African American ✢✢✢	-0.223***	0.800
Asian ✢✢✢	-0.148**	0.863
Other Race ✢✢✢	-0.110	0.896
Race Unknown ✢✢✢	0.010	1.010
Semester GPA in 2s	-0.421***	0.657
Semester GPA in 1s	-1.101***	0.332
Semester GPA less than 1	-1.994***	0.136
Tau	6.237	
Intercept Reliability	0.735	

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

✢ Social = ref., ✢✢ First year = ref., ✢✢✢ Caucasian/White = ref.

Table L2: **Level-one descriptive statistics: Hypothesis two excluding students without SATs.**

VARIABLE NAME	*N*	*MEAN*	*SD*	*MIN.*	*MAX.*
SET SUBMISSION	119,190	0.49	0.50	0.00	1.00
GRADEA	119,190	0.33	0.47	0.00	1.00
GRADEB	119,190	0.29	0.45	0.00	1.00
GRADEC	119,190	0.13	0.34	0.00	1.00
GRADED	119,190	0.04	0.19	0.00	1.00
GRADEFU	119,190	0.04	0.20	0.00	1.00
GRADENG	119,190	0.11	0.32	0.00	1.00
GRADES	119,190	0.05	0.22	0.00	1.00
CRSGRDA	119,190	0.15	0.35	0.00	1.00
CRSGRDB	119,190	0.51	0.50	0.00	1.00
CRSCUNDE	119,190	0.23	0.42	0.00	1.00

Table L3: **Level-two descriptive statistics: Hypothesis two excluding students without SATs.**

VARIABLE NAME	*N*	*MEAN*	*SD*	*MIN.*	*MAX.*
SAT	19,612	1178.64	134.41	600.00	1600.00
TOTALCH	19,612	60.18	42.29	0.00	355.00
NCSUCH	19,612	58.16	40.69	0.00	301.00
SEMCH	19,612	14.78	2.29	0.00	23.00
FEMALEYN	19,612	0.44	0.50	0.00	1.00
TRNSFRYN	19,612	0.06	0.24	0.00	1.00
RESYN	19,612	0.92	0.26	0.00	1.00
UNDMAJ	19,612	0.13	0.33	0.00	1.00
AFRAMER	19,612	0.09	0.28	0.00	1.00
ASIAN	19,612	0.05	0.23	0.00	1.00
OTHRRACE	19,612	0.04	0.19	0.00	1.00
RACEUNRP	19,612	0.04	0.20	0.00	1.00
TRADAGE	19,612	0.99	0.11	0.00	1.00
PARTYN	19,612	0.04	0.19	0.00	1.00
DBMAJYN	19,612	0.05	0.22	0.00	1.00
CAMPUSYN	19,612	0.36	0.48	0.00	1.00
ATHLYN	19,612	0.03	0.16	0.00	1.00
INTL	19,612	0.01	0.09	0.00	1.00
SOPH	19,612	0.24	0.43	0.00	1.00
JUNIOR	19,612	0.23	0.42	0.00	1.00
SENIOR	19,612	0.25	0.43	0.00	1.00
GPAIN2S	19,612	0.42	0.49	0.00	1.00
GPAIN1S	19,612	0.06	0.23	0.00	1.00
GPALESS1	19,612	0.01	0.09	0.00	1.00
ARTISTIC	19,612	0.03	0.16	0.00	1.00
CONVENT	19,612	0.02	0.13	0.00	1.00
ENTERP	19,612	0.15	0.36	0.00	1.00
INVEST	19,612	0.47	0.50	0.00	1.00
REALIST	19,612	0.19	0.39	0.00	1.00
UNDECMAJ	19,612	0.07	0.25	0.00	1.00

Table L4: **Fully controlled model: Hypothesis two excluding students without SATs**

Variable	***Controlled Model***	
	Coeff.	*Odds Ratio*
Level One *(Grade A = ref. group)*	-0.066***	0.937
Student Grade = B	0.003	1.003
Student Grade = C	-0.029*	0.972
Student Grade = D	-0.056**	0.946
Student Grade = F/U	-0.412***	0.662
Student Grade = Satisfactory (S)	-0.348***	0.706
No Grade	-0.476***	0.621
Level Two	-0.091***	0.913
Artistic✝	0.080	1.084
Conventional✝	0.184	1.202
Enterprising✝	-0.011	0.989
Investigative✝	0.062	1.064
Realistic✝	0.252***	1.286
FYC/Transitions✝	0.137*	1.147
Sophomore✝✝	-0.305***	0.737
Junior✝✝	-0.299***	0.742
Senior✝✝	-0.151	0.859
NCSU CH	-0.000	1.000
Semester CH	0.024**	1.025
Female	0.315***	1.370
On campus Housing	0.252***	1.287
Athlete	-0.419***	0.657
International Student	0.079	1.082
Transfer Student	0.040	1.041
NC Resident	0.039	1.040
Traditional Age (< 25 years)	-0.039	0.962
Part-time	-0.166	0.847
African American ✝✝✝	-0.377***	0.686
Asian ✝✝✝	-0.168**	0.846
Other Race ✝✝✝	-0.172**	0.842
Race Unknown ✝✝✝	-0.028	0.972
Semester GPA in 2s	-0.446***	0.640
Semester GPA in 1s	-1.156***	0.315
Semester GPA less than 1	-2.083***	0.125
SAT score	-0.001***	0.999
Tau	6.182	
Intercept Reliability	0.736	

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

✝ Social = ref., ✝✝ First year = ref., ✝✝✝ Caucasian/White = ref.

Appendix M: Average Course Grade Models

Average Course Grade Models: Omitted from Hypothesis Two

Level-two predictors in this model remained the same as the Grades Earned in Course model. However, at level-one, instead of taking the student's individual grade into account, the average grade for the entire class became the level-one variable. Perhaps it may be acceptable to assume that little would change, however, some differences were noticeable. Moreover, the differences in these models (Table M1) may be more practically applicable to addressing SET nonresponse as a whole. An ungraded course was the reference group. That is, these were courses that could raise or lower a student's GPA. If the course did not award GPA points to the student (e.g. there was no grade for the course or it was a pass/fail course), it was part of this category.

Overall, if the course outcome awarded GPA points, students were much more likely to complete course evaluations than the mean. At level-two, students with GPAs below 3.0 were less likely to respond. Again SAT score emerged as statistically significant, but statistical significance does not always mean practical significance.

***Table M1*: Average course grade: Comparisons with and without SAT scores.**

Variable	*All Students*		*Only Students With SAT*	
	Coeff.	*Odds Ratio*	*Coeff.*	*Odds Ratio*
Level One *(Ungraded Course = ref.)*	-0.056***	0.946	-0.072***	0.931
Average Course Grade = A	0.316***	1.372	0.303***	1.354
Av. Course Grade = B	0.415***	1.515	0.416***	1.515
Av. Course Grade = C & Below	0.312***	1.365	0.310***	1.364
Level Two *(3.0-4.0 GPA = ref.)*	-0.0633***	0.939	-0.080***	0.923
Semester GPA in 2s	-0.534***	0.586	-0.564***	0.569
Semester GPA in 1s	-1.185***	0.306	-1.255***	0.285
Semester GPA less than 1	-2.065***	0.127	-2.123***	0.120
SAT score	-----	-----	-0.001***	0.999
Interactions (* Grade A)	0.322***	1.380	0.309***	1.361
Semester GPA in 2s	-0.009	0.991	-0.005	0.995
Semester GPA in 1s	0.037	1.038	0.061	1.062
Semester GPA less than 1	0.082	1.086	0.017	1.017
SAT score	-----	-----	0.000	0.999
Interactions (* Grade B)	0.420***	1.522	0.420***	1.521
Semester GPA in 2s	-0.096***	0.908	-0.104***	0.901
Semester GPA in 1s	-0.112	0.894	-0.111	0.895
Semester GPA less than 1	-0.091	0.913	-0.088	0.915
SAT score	-----	-----	-0.000	1.000
Interactions (* Grade ≤ C)	0.323***	1.381	0.320***	1.377
Semester GPA in 2s	-0.096**	0.909	-0.096***	0.908
Semester GPA in 1s	-0.053	0.948	-0.039	0.962
Semester GPA less than 1	-0.070	0.932	-0.185	0.831
SAT score	-----	-----	-0.000	1.000
Tau	6.378		6.420	
Intercept Reliability	0.745		0.750	

*** p<0.001, **p<0.01, *p<0.05

Summary of Interactions With Grades Earned

For several variables, interactions (Table M2) somewhat controlled for the statistical significance in non-interactive models (Tables 4.5 and 4.6). No matter what the semester GPA was, if the student received a B, NG, or S in the class, they did not differ from the average likelihood of the higher semester GPA group. For C, D, and F grades, students with a GPA between 1.0 and 1.99 were more likely to respond, with no statistically significant changes among other variable interactions.

Table M2: **Achievement models: Grade earned in course.**

Variable	*All Students*		*Only Students With SAT*	
	Coeff.	*Odds ratio*	*Coeff.*	*Odds ratio*
Level One *(Grade A = ref)*	-0.056***	0.945	-0.072***	0.930
Student Grade = B	-0.029*	0.971	-0.021*	0.980
Student Grade = C	-0.093***	0.912	-0.080***	0.923
Student Grade = D	-0.149***	0.861	-0.136***	0.873
Student Grade = F/U	-0.350***	0.705	-0.520***	0.594
Student Grade = Satisfactory (S)	-0.531***	0.588	-0.343***	0.709
No Grade	-0.505***	0.603	-0.489***	0.613
Level Two *(GPA 3.0-4.0 = ref)*	-0.072***	0.930	-0.089***	0.914
Semester GPA in 2s	-0.498***	0.608	-0.533***	0.587
Semester GPA in 1s	-1.135***	0.321	-1.224***	0.294
Semester GPA less than 1	-1.903***	0.149	-1.979***	0.138
SAT score	-----	-----	-0.001***	1.000
Interactions (* Grade B)	**-0.001**	**0.999**	**0.007**	**1.007**
Semester GPA in 2s	-0.011	0.989	-0.014	0.986
Semester GPA in 1s	0.053	1.055	0.073	1.076
Semester GPA less than 1	0.507	1.661	0.482	1.619
SAT score	-----	-----	-0.001	1.000
Interactions (* Grade C)	**-0.050*** **	**0.952**	**-0.041*** **	**0.960**
Semester GPA in 2s	0.048	1.049	0.052	1.054
Semester GPA in 1s	0.155	1.168	0.199	1.220
Semester GPA less than 1	0.289	1.336	0.135	1.144
SAT score	-----	-----	-0.000	1.000
Interactions (* Grade D)	**-0.108*** **	**0.897**	**-0.093*** **	**0.911**
Semester GPA in 2s	0.085	1.089	0.069	1.071
Semester GPA in 1s	0.229**	1.257	0.252**	1.286
Semester GPA less than 1	0.620	1.859	0.614	1.848
SAT score	-----	-----	-0.000	1.000
Interactions (* Grade F/U)	**-0.429*** **	**0.651**	**-0.401*** **	**0.670**
Semester GPA in 2s	-0.051	0.951	-0.137	0.872
Semester GPA in 1s	0.197	1.217	0.159	1.172
Semester GPA less than 1	0.289	1.335	0.160	1.173
SAT score	-----	-----	-0.001	1.000
Interactions (* No Grade)	**-0.477*** **	**0.620**	**-0.464*** **	**0.629**
Semester GPA in 2s	-0.029	0.972	-0.016	0.984
Semester GPA in 1s	0.015	1.015	0.049	1.051
Semester GPA less than 1	0.067	1.069	0.087	1.091
SAT score	-----	-----	0.000	1.000
Interactions (* Grade S)	**-0.328*** **	**0.720**	**-0.320*** **	**0.726**
Semester GPA in 2s	0.064	1.066	0.066	1.069
Semester GPA in 1s	0.111	1.117	0.085	1.088
Semester GPA less than 1	0.055	1.057	0.036	1.036
SAT score	-----	-----	-0.000	1.000
Tau	6.354		6.397	
Intercept Reliability	0.743		0.747	

*** p<0.001, **p<0.01, *p<0.05

Appendix N: Hypothesis Three Survey Fatigue and Opportunity Costs

Table N1. **Survey fatigue: Fully controlled model.**

Variable	***Controlled Model***	
	Coeff.	*Odds Ratio*
Intercept	-0.063***	0.939
NCSU CH	-0.002*	0.998
SETs administered	0.016*	1.016
On campus Housing	0.231***	1.260
Athlete	-0.382***	0.682
International Student	0.231	1.259
Class Rank *(First-year = ref.)*		
Sophomore	-0.247***	0.782
Junior	-0.183**	0.832
Senior	-0.013	0.987
Academic Environment *(Social = ref.)*		
Artistic	0.058	1.059
Conventional	0.087	1.091
Enterprising	-0.028	0.972
Investigative	0.000	1.000
Realistic	0.183***	1.200
FYC/Transitions	0.132*	1.142
Race *(White/Caucasian = ref.)*		
African American	-0.226***	0.798
Asian	-0.154**	0.857
Other Race	-0.111	0.895
Race Unknown	0.007	1.007
Semester GPA *(3.0-4.0 = ref.)*		
Semester GPA in 2s	-0.446***	0.640
Semester GPA in 1s	-1.198***	0.302
Semester GPA less than 1	-2.230**	0.108
Transfer Student	0.180***	1.197
NC Resident	0.019	1.019
Traditional Age (< 25 years)	-0.362***	0.696
Female	0.351***	1.421
Part-time	-0.041	0.960
Tau	6.059	
Intercept Reliability	0.735	

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

***Table N2*. Models with squared terms.**

		Credit Hours		***Controlled***	
Variable		*Coeff.*	*Odds Ratio*	*Coeff.*	*Odds Ratio*
	Intercept	-0.054***	0.947	-0.063*	0.939
Total CH		-0.003***	0.998	-0.003***	0.997
	Total CH (Squared)	-0.000**	1.000	-0.000*	1.000
Semester CH		0.035***	1.036	0.015*	1.016
	Semester CH (Squared)	0.001***	1.001	0.000*	1.000
	Tau	6.369		6.059	
	Intercept Reliability	0.752		0.735	

*** p<0.001, **p<0.01, *p<0.05

Appendix O: Hypothesis Four: Excluding Undecided Majors

Table O1: **Descriptive statistics: Academic environment excluding undecided majors.**

VARIABLE	*N*	*MEAN*	*SD*	*RANGE*	
				MIN	*MAX*
SETs	125,908	-----	-----	-----	-----
STUDENTS	21,226	-----	-----	-----	-----
RESPONSE	21,226	0.49	0.50	0.00	1.00
CRS IN MAJOR	21,226	0.27	0.45	0.00	1.00
ARTISTIC*	21,226	0.03	0.17	0.00	1.00
CONVENT*	21,226	0.02	0.14	0.00	1.00
ENTERPRI *	21,226	0.17	0.37	0.00	1.00
INVESTIG*	21,226	0.50	0.50	0.00	1.00
REALISTIC*	21,226	0.20	0.40	0.00	1.00

*Social = reference group.

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$\text{Prob}(\text{URESPYN}=1 \mid \pi) = \varphi$$

$$\text{Log}[\varphi/(1 - \varphi)] = \eta$$

$$\eta = \pi_0 + \pi_1(\boldsymbol{CRSNMAJC})$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\begin{aligned}
\pi_0 = {} & \beta_{00} + \beta_{01}(\boldsymbol{NCSUCH}) + \beta_{02}(\boldsymbol{SEMCH}) + \beta_{03}(\boldsymbol{CAMPUSYN}) + \beta_{04}(\boldsymbol{ATHLYN}) + \\
& \beta_{05}(\boldsymbol{INTL}) + \beta_{06}(\boldsymbol{SOPH}) + \beta_{07}(\boldsymbol{JUNIOR}) + \beta_{08}(\boldsymbol{SENIOR}) + \beta_{09}(\boldsymbol{GPAIN2S}) + \\
& \beta_{010}(\boldsymbol{GPAIN1S}) + \beta_{011}(\boldsymbol{GPALESS1}) + \beta_{012}(\boldsymbol{TRNSFRYN}) + \beta_{013}(\boldsymbol{RESYN}) + \\
& \beta_{014}(\boldsymbol{AFRAMER}) + \beta_{015}(\boldsymbol{ASIAN}) + \beta_{016}(\boldsymbol{OTHRRACE}) + \beta_{017}(\boldsymbol{RACEUNRP}) + \\
& \beta_{018}(\boldsymbol{TRADAGE}) + \beta_{019}(\boldsymbol{PARTYN}) + \beta_{020}(\boldsymbol{FEMALEYN}) + \beta_{021}(\boldsymbol{ARTISTIC}) + \\
& \beta_{022}(\boldsymbol{CONVENT}) + \beta_{023}(\boldsymbol{ENTERPRI}) + \beta_{024}(\boldsymbol{INVESTIG}) + \beta_{025}(\boldsymbol{REALISTI}) + r_0
\end{aligned}$$

$$\begin{aligned}
\pi_1 = {} & \beta_{10} + \beta_{11}(\boldsymbol{NCSUCH}) + \beta_{12}(\boldsymbol{SEMCH}) + \beta_{13}(\boldsymbol{CAMPUSYN}) + \beta_{14}(\boldsymbol{ATHLYN}) + \\
& \beta_{15}(\boldsymbol{INTL}) + \beta_{16}(\boldsymbol{SOPH}) + \beta_{17}(\boldsymbol{JUNIOR}) + \beta_{18}(\boldsymbol{SENIOR}) + \beta_{19}(\boldsymbol{GPAIN2S}) + \\
& \beta_{110}(\boldsymbol{GPAIN1S}) + \beta_{111}(\boldsymbol{GPALESS1}) + \beta_{112}(\boldsymbol{TRNSFRYN}) + \beta_{113}(\boldsymbol{RESYN}) + \\
& \beta_{114}(\boldsymbol{AFRAMER}) + \beta_{115}(\boldsymbol{ASIAN}) + \beta_{116}(\boldsymbol{OTHRRACE}) + \beta_{117}(\boldsymbol{RACEUNRP}) + \\
& \beta_{118}(\boldsymbol{TRADAGE}) + \beta_{119}(\boldsymbol{PARTYN}) + \beta_{120}(\boldsymbol{FEMALEYN}) + \beta_{121}(\boldsymbol{ARTISTIC}) + \\
& \beta_{122}(\boldsymbol{CONVENT}) + \beta_{123}(\boldsymbol{ENTERPRI}) + \beta_{124}(\boldsymbol{INVESTIG}) + \beta_{125}(\boldsymbol{REALISTI}) + r_1
\end{aligned}$$

Figure O1: **Equations for models in hypothesis four.**

$$
\begin{aligned}
\eta = {} & \beta_{00} + \beta_{01}*NCSUCH + \beta_{02}*SEMCH + \beta_{03}*CAMPUSYN + \beta_{04}*ATHLYN + \\
& \beta_{05}*INTL + \beta_{06}*SOPH + \beta_{07}*JUNIOR + \beta_{08}*SENIOR + \beta_{09}*GPAIN2S + \\
& \beta_{010}*GPAIN1S + \beta_{011}*GPALESS1 + \beta_{012}*TRNSFRYN + \beta_{013}*RESYN + \\
& \beta_{014}*AFRAMER + \beta_{015}*ASIAN + \beta_{016}*OTHRRACE + \beta_{017}*RACEUNRP + \\
& \beta_{018}*TRADAGE + \beta_{019}*PARTYN + \beta_{020}*FEMALEYN + \beta_{021}*ARTISTIC + \\
& \beta_{022}*CONVENT + \beta_{023}*ENTERPRI + \beta_{024}*INVESTIG + \beta_{025}*REALISTI + \\
& \beta_{10}*CRSNMAJC + \beta_{11}*NCSUCH*CRSNMAJC + \beta_{12}*SEMCH*CRSNMAJC + \\
& \beta_{13}*CAMPUSYN*CRSNMAJC + \beta_{14}*ATHLYN*CRSNMAJC + \\
& \beta_{15}*INTL*CRSNMAJC + \beta_{16}*SOPH*CRSNMAJC + \beta_{17}*JUNIOR*CRSNMAJC + \\
& \beta_{18}*SENIOR*CRSNMAJC + \beta_{19}*GPAIN2S*CRSNMAJC + \\
& \beta_{110}*GPAIN1S*CRSNMAJC + \beta_{111}*GPALESS1*CRSNMAJC + \\
& \beta_{112}*TRNSFRYN*CRSNMAJC + \beta_{113}*RESYN*CRSNMAJC + \\
& \beta_{114}*AFRAMER*CRSNMAJC + \beta_{115}*ASIAN*CRSNMAJC + \\
& \beta_{116}*OTHRRACE*CRSNMAJC + \beta_{117}*RACEUNRP*CRSNMAJC + \\
& \beta_{118}*TRADAGE*CRSNMAJC + \beta_{119}*PARTYN*CRSNMAJC + \\
& \beta_{120}*FEMALEYN*CRSNMAJC + \beta_{121}*ARTISTIC*CRSNMAJC + \\
& \beta_{122}*CONVENT*CRSNMAJC + \beta_{123}*ENTERPRI*CRSNMAJC + \\
& \beta_{124}*INVESTIG*CRSNMAJC + \beta_{125}*REALISTI*CRSNMAJC + r_0
\end{aligned}
$$

Figure O2: **Mixed model equation: See Figure O1**

Table O2: **Fully controlled model: Course location.**

Variable	***Controlled Model***	
	Coeff.	*Odds Ratio*
Level One	-0.053***	0.948
Course in Dept. of Student's Major	0.059***	1.061
Level Two	-0.069***	0.933
Semester CH	0.015*	1.015
Female	0.348***	1.416
Artistic✝	0.054	1.056
Conventional✝	0.093	1.097
Enterprising✝	-0.027	0.973
Investigative✝	0.015	1.015
Realistic✝	0.188***	1.207
Sophomore✝✝	-0.212***	0.809
Junior✝✝	-0.176*	0.839
Senior✝✝	-0.020	0.980
NCSU CH	-0.002	0.998
On campus Housing	0.214***	1.239
Athlete	-0.277***	0.758
International Student	0.235	1.265
Transfer Student	0.178***	1.195
NC Resident	0.024	1.025
Traditional Age (< 25 years)	-0.356***	0.701
Part-time	-0.041	0.960
African American ✝✝✝	-0.211***	0.810
Asian ✝✝✝	-0.157**	0.855
Other Race ✝✝✝	-0.112	0.894
Race Unknown ✝✝✝	0.001	1.001
Semester GPA in 2s	-0.438***	0.646
Semester GPA in 1s	-1.216***	0.296
Semester GPA less than 1	-2.232***	0.107

Continued on next page.

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

✝ Social = ref., ✝✝ First year = ref., ✝✝✝ Caucasian/White = ref.

***Table O2*: (continued).**

Variable	*Controlled Model*	
	Coeff.	*Odds Ratio*
Interactions (* Crs in Major Dept.)	0.057***	1.058
Semester CH	-0.001	0.999
Female	-0.027	0.973
Artistic✝	0.239***	1.269
Conventional✝	0.215**	1.239
Enterprising✝	0.155**	1.168
Investigative✝	0.138**	1.148
Realistic✝	0.161**	1.175
Sophomore✝✝	0.164***	1.178
Junior✝✝	0.042	1.043
Senior✝✝	0.028	1.029
NCSU CH	0.000	1.000
On campus Housing	0.008	1.008
Athlete	0.131*	1.140
International Student	0.061	1.063
Transfer Student	-0.066*	0.936
NC Resident	0.020	1.021
Traditional Age (Age < 25 years)	-0.030	0.971
Part-time	-0.015	0.985
African American ✝✝✝	0.021	1.021
Asian ✝✝✝	-0.011	0.989
Other Race ✝✝✝	0.008	1.008
Race Unknown ✝✝✝	0.011	1.011
Semester GPA in 2s	0.014	1.014
Semester GPA in 1s	-0.077	0.926
Semester GPA less than 1	-0.173	0.842
Tau	6.182	
Intercept Reliability	0.736	

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

✝ Social = ref., ✝✝ First year = ref., ✝✝✝ Caucasian/White = ref.

Appendix P: Hypothesis Five: Fully Controlled Model With Interactions

Table P1. **Fully controlled model: Grade earned in course.**

Variable	***Controlled Model***	
	Coeff.	*Odds Ratio*
Level One *(Grade Earned: A,B,C,S = ref.)*	-0.055***	0.947
Grade = D and F	-0.261***	0.770
No Grade	-0.453***	0.636
Level Two	-0.073***	0.930
Artistic✝	0.051	1.052
Conventional✝	0.088	1.092
Enterprising✝	-0.049	0.952
Investigative✝	0.031	1.032
Realistic✝	0.213***	1.238
FYC✝	0.131**	1.140
Sophomore✝✝	-0.228***	0.796
Junior✝✝	-0.180*	0.835
Senior✝✝	-0.021	0.979
NCSU CH	-0.002*	0.998
Semester CH	0.016*	1.017
Female	0.352***	1.422
On campus Housing	0.227***	1.255
Athlete	-0.391***	0.677
International Student	0.228	1.256
Transfer Student	0.180***	1.200
NC Resident	0.017	1.017
Traditional Age (Age < 25 years)	-0.374***	0.688
Part-time	-0.023	0.978
African American ✝✝✝	-0.222***	0.801
Asian ✝✝✝	-0.152**	0.859
Other Race ✝✝✝	-0.107	0.898
Race Unknown ✝✝✝	0.007	1.007
Semester GPA in 2s	-0.419***	0.657
Semester GPA in 1s	-1.127***	0.324
Semester GPA less than 1	-2.034***	0.131

Continued on the next page.

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

✝ Social = ref., ✝✝ First year = ref., ✝✝✝ Caucasian/White = ref.

Table P1: **(continued).**

Variable	***Controlled Model***	
	Coeff.	*Odds Ratio*
Interactions (* D/F Grades.)	-0.254***	0.776
Artistic✝	-0.158	0.854
Conventional✝	0.203	1.225
Enterprising✝	-0.072	0.931
Investigative✝	0.056	1.057
Realistic✝	0.107	1.112
FYC✝	0.008	1.008
Sophomore✝✝	-0.017	0.983
Junior✝✝	0.085	1.089
Senior✝✝	0.075	1.077
NCSU CH	-0.002	0.998
Semester CH	0.029***	1.030
Female	0.006	1.006
On campus Housing	0.102**	1.107
Athlete	0.084	1.088
International Student	0.054	1.055
Transfer Student	0.171***	1.142
NC Resident	-0.108	0.898
Traditional Age (Age < 25 years)	0.133	1.366
Part-time	-0.347**	0.707
African American ✝✝✝	-0.037	0.964
Asian ✝✝✝	-0.017	0.983
Other Race ✝✝✝	-0.012	0.989
Race Unknown ✝✝✝	-0.018	0.982
Semester GPA in 2s	0.036	1.037
Semester GPA in 1s	0.074	1.077
Semester GPA less than 1	-0.031	0.969

Continued on the next page.

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

✝ Social = ref., ✝✝ First year = ref., ✝✝✝ Caucasian/White = ref.

Table P1: **(continued.)**

Variable	***Controlled Model***	
	Coeff.	*Odds Ratio*
Interactions (* No Grade)	-0.501***	0.606
Artistic✝	0.076	1.079
Conventional✝	0.080	1.083
Enterprising✝	-0.033	0.968
Investigative✝	0.252***	1.288
Realistic✝	0.196**	1.218
FYC✝	0.046	1.047
Sophomore✝✝	-0.092	0.915
Junior✝✝	0.107	1.113
Senior✝✝	0.059	1.060
NCSU CH	-0.001	0.999
Semester CH	-0.004	0.996
Female	-0.010	0.989
On campus Housing	-0.012	0.988
Athlete	-0.184	0.832
International Student	-0.046	0.955
Transfer Student	0.000	1.000
NC Resident	-0.047	0.954
Traditional Age (Age < 25 years)	0.138	1.148
Part-time	-0.354**	0.705
African American ✝✝✝	0.047	1.048
Asian ✝✝✝	0.019	1.020
Other Race ✝✝✝	-0.035	0.966
Race Unknown ✝✝✝	-0.037	0.964
Semester GPA in 2s	-0.020	0.976
Semester GPA in 1s	-0.050	0.951
Semester GPA less than 1	-0.130	0.873
Tau	6.190	
Intercept Reliability	0.734	

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

✝ Social = ref., ✝✝ First year = ref., ✝✝✝ Caucasian/White = ref.

Appendix Q: Hypothesis Five Expanded Uncontrolled Model

This is the expanded version of Table 4.12 and an overview of the outcomes. Without interactions all grades are statistically significant in predicting a decrease in the likelihood of response from the mean. A fully controlled model displayed errors due to a large file size, so its outcomes are not below.

In the uncontrolled model without interactions (grades and environment), only the enterprising majors were statistically significant ($p < .05$), less likely to respond. However, for the between-students analyses, grades interacted with the types and tendencies to respond changed. For example, an enterprising student obtaining a B in the course displayed no difference from the mean in the likelihood to complete that course's SET. Realistic students were more likely to respond (with statistical significance) than the mean when their grade was a B, C, D, or if they did not receive a grade. Though they were statistically different without interactions, with grades interacting enterprising students were either not statistically different or did not decrease from the mean when compared to social students with C grades. Investigative students were statistically significant with grades of B (more likely to respond) or if no grade was given at all (less likely). Artistic students were less likely to respond when they received Ds and Fs, more likely for satisfactory grades. Finally, First Year College and Transitions students were less likely to respond to an SET for a course in which they received an S.

Table Q1: **SET salience: Uncontrolled model only.**

Variable	*Grades and Environment*	
	Coeff.	*Odds Ratio*
Level One	-0.055***	0.946
Grade Earned in Course (*A = ref.)*		
Grade = B	-0.030***	0.971
Grade = C	-0.093***	0.911
Grade = D and F	-0.323***	0.705
Grade = S	-0.349***	0.724
Grade Not Awarded	-0.504***	0.604
Level Two	-0.060***	0.942
Academic Environment (*Social = ref)*		
Artistic	0.041	1.041
Conventional	0.031	1.031
Enterprising	-0.169*	0.844
Investigative	-0.049	0.952
Realistic	0.061	1.063
FYC/Transitions	0.011	1.011
Interactions (* Grade B)	-0.032***	0.969
Artistic	0.018	1.018
Conventional	0.051	1.053
Enterprising	0.041	1.042
Investigative	0.068	1.071
Realistic	0.099**	1.104
FYC/Transitions	-0.021	0.979
Interactions (* Grade C)	-0.099***	0.905
Artistic	-0.152	0.859
Conventional	-0.108	0.897
Enterprising	-0.099	0.906
Investigative	0.038	1.038
Realistic	0.152***	1.165
FYC/Transitions	-0.020	0.980
Interactions (* Grade D & F)	-0.340***	0.712
Artistic	-0.266**	0.766
Conventional	0.106	1.112
Enterprising	-0.098	0.906
Investigative	0.084	1.087
Realistic	0.190***	1.209
FYC/Transitions	0.031	1.032
Interactions (* Grade S/P)	-0.343***	0.709
Artistic	0.275**	1.317
Conventional	-0.023	0.978
Enterprising	0.044	1.044
Investigative	-0.028	0.972
Realistic	0.010	1.010
FYC/Transitions	-0.177	0.837

Table Q1: **(continued.)**

Variable	***Grades and Environment***	
	Coeff.	*Odds Ratio*
Interactions (* No Grade)	-0.541***	0.582
Artistic	0.023	1.023
Conventional	0.068	1.071
Enterprising	-0.032	0.969
Investigative	0.270***	1.310
Realistic	0.253***	1.288
FYC/Transitions	0.014	1.014
Tau	6.443	
Intercept Reliability	0.751	

*** p<0.001, **p<0.01, *p<0.05