

Spring 2013

The effects of item and respondent characteristics on midpoint response option endorsement: A mixed-methods study

Kimberly Rebecca Marsh
James Madison University

Follow this and additional works at: <https://commons.lib.jmu.edu/diss201019>

 Part of the [Psychology Commons](#)

Recommended Citation

Marsh, Kimberly Rebecca, "The effects of item and respondent characteristics on midpoint response option endorsement: A mixed-methods study" (2013). *Dissertations*. 72.
<https://commons.lib.jmu.edu/diss201019/72>

This Dissertation is brought to you for free and open access by the The Graduate School at JMU Scholarly Commons. It has been accepted for inclusion in Dissertations by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.

The Effects of Item and Respondent Characteristics on Midpoint Response Option

Endorsement: A Mixed-Methods Study

Kimberly R. Marsh

A dissertation submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

Department of Graduate Psychology

May, 2013

Acknowledgements

I would like to begin by thanking my advisor, Dr. Dena Pastor for all of her invaluable time, guidance, and support. I am extremely grateful to have benefited from her superior teaching, mentorship, and scholarship over the course of the past five years and look forward to continued collaboration in the future. Thank you, Dena, for being an exceptional personal and professional role model.

Next, I would like to express my gratitude to my other committee members: Dr. Debbi Bandalos, Dr. Christine DeMars, and Dr. Megan France, who have been incredibly generous with their expertise, insight, and support throughout the entire dissertation process. I simply could not ask for a better collection of mentors contributing to this line of research and my development as a researcher.

I also wish to acknowledge and thank each member of the Center for Assessment and Research Studies faculty and staff for their contributions to my personal and professional development over the past five years. I wish to specifically thank Dr. Sara Finney, Dr. Donna Sundre, Dr. Robin Anderson, and Dr. Keston Fulcher. Thank you each for your dedication to teaching and mentorship, allowing me the opportunity and honor to collaborate on research, inspiring me to be a better person and professional, and providing your unwavering support and encouragement over the years.

Finally, I would like thank my friends and family who have been and will always be the reason I am able to achieve any goal. I am especially grateful to my husband, Chris; thank you for your unconditional love and encouragement. I could not have done this without you and of course the other two members of our little family, our fur babies, Daisy and Delilah.

Table of Contents

Acknowledgements.....	ii
List of Tables	viii
List of Figures.....	ix
Abstract.....	x
I. Introduction	1
Chapter Overview	1
Noncognitive Assessment in a Higher Education Context.....	1
Response Error and Validity of Respondent Scores.....	4
The midpoint response option.....	6
Survey Instruments versus Noncognitive Assessments.....	7
Response processes.....	8
Response styles.....	9
Item and respondent characteristics.....	10
Purpose of the Current Study.....	11
II. Review of the Literature.....	15
Chapter Overview	15
Proper Use versus Abuse of the Midpoint Response Option in Practice.....	15
Midpoint Response Option Abuse in Assessment and Measurement Practice..	20
Midpoint response option label and midpoint abuse	22
Midpoint Response Style (MRS) Research	26
Item Characteristics and MRS	28
Respondent Characteristics and MRS.....	32

Both Item and Respondent Characteristics and MRS.....	34
Research Questions.....	42
Study 1	43
Study 2	46
III. Methods.....	48
Chapter Overview	48
Study 1	49
Participants and Procedure.....	49
Measures	50
Sociocultural Domain Assessment Version 7 (SDA-7).....	50
Student Opinion Scale (SOS).....	52
Item Characteristics	53
Item set location.....	53
Midpoint response option label.....	53
Respondent Characteristics.....	54
Student self-reported effort.....	54
SAT verbal score (SATV)	54
Data Analysis	54
Model 1 – Unconditional Model.....	55
Model 2 – Item set location, midpoint response option label, and their interaction effects.....	56
Model 3 – Item set location, midpoint response option label, respondent self-reported effort, and their interaction effects	56

Model 4 – Item set location, midpoint response option label, respondent SAT verbal score, and their interaction effects.....	57
Interpretation of Results.....	58
Software	58
Study 2	58
Participants and Procedure.....	58
Measures	60
Data Analysis	61
IV. Results.....	63
Chapter Overview	63
Study 1	63
Data Cleaning.....	63
Descriptive Statistics.....	63
Model 1 – Unconditional Model.....	65
Model 2 – Item set location, midpoint response option label, and their interaction effects.....	66
Model 3 – Item set location, midpoint response option label, respondent self-reported effort, and their interaction effects	67
Model 4 – Item set location, midpoint response option label, respondent SAT verbal score, and their interaction effects.....	69
Models 5 and 6 – Item set location, midpoint response option label, respondent self-reported effort, SAT verbal score, and their interaction effects	71

Study 1 Summary.....	72
Research Question 1	73
Research Question 2	74
Research Question 3	75
Research Question 4	75
Research Question 5	75
Study 2	76
Study 2 Research Question 1: Think-aloud Results	76
Ambivalence	77
Indecision.....	78
Item statement clarity.....	80
Study 2 Research Question 2: Structured Interview Question Results.....	81
<i>Neutral</i>	84
<i>Neither Agree nor Disagree</i>	84
<i>Undecided</i>	85
<i>No opinion</i>	85
Study 2 Summary.....	86
Research Question 1	86
Research Question 2	87
V. Discussion	89
Integration and Summary of Findings from Study 1 and Study 2	90
Implications of Findings for Assessment Practice.....	93
Limitations of the Current Study and Directions for Future Research	98

General Conclusion.....	102
Footnotes.....	104
Appendices.....	105
References.....	139

List of Tables

Table 1: Fall 2012 Assessment Day Testing Session Configurations	146
Table 2: Study 1 SDA-7 SoI and M-GUDS Forms	147
Table 3: Study 1 Variable Intercorrelations and Descriptive Statistics	148
Table 4: Study 1 Variable Descriptive Statistics by SDA-7 Form and Two-way ANOVA Results.....	149
Table 5: Study 1 Model 2 Results.....	150
Table 6: Study 1 Model 3 Results.....	150
Table 7: Study 1 Model 4 Results.....	151
Table 8: Study 1 Exploratory Model 5 Results.....	152
Table 9: Study 1 Exploratory Model 6 Results.....	153
Table 10: Study 1 Fit Statistics, τ_{00} for Models 2-6, and Likelihood Ratio Test Results Comparing Models 2, 3, 4, and 6 to Model 5	154
Table 11: Study 2 Think-aloud Coding Scheme.....	155
Table 12: Study 2 Think-aloud Midpoint Endorsement Summary.....	156

List of Figures

Figure 1: Histogram of average midpoint response endorsement across items for the 1,826 respondents	157
Figure 2: Predicted probability of midpoint response (MR) endorsement across model 2 predictor combinations.....	158
Figure 3: Predicted probability of midpoint response (MR) endorsement across model 3 predictor combinations.....	159
Figure 4: Predicted probability of midpoint response (MR) endorsement across model 4 predictor combinations.....	160
Figure 5: Predicted probability of midpoint response (MR) endorsement for high SAT verbal respondents across model 6 predictor combinations	161
Figure 6: Predicted probability of midpoint response (MR) endorsement for low SAT verbal respondents across model 6 predictor combinations	161

Abstract

As the demand for accountability and transparency in higher education has increased, so too has the call for direct assessment of student learning outcomes. Accompanying this increase of knowledge-based, cognitive assessments administered in a higher education context is an increased emphasis on assessing various noncognitive aspects of student growth and development over the course of their college career. Noncognitive outcomes are most often evaluated via self-report instruments associated with Likert-type response scales, posing unique challenges for researchers and assessment practitioners hoping to draw valid conclusions based upon this data. One long-debated characteristic of such assessments is the midpoint response option. More specifically, prior research suggests that respondents may be more or less likely to endorse the midpoint response option under different measurement and respondent dispositional conditions thus introducing construct-irrelevant variance within respondent scores. The current study expanded upon previous work to examine the effects of various item and respondent characteristics on endorsement and conceptualization of the midpoint response option in a noncognitive assessment context.

A mixed-methods approach was employed in order to fully address research questions associated with two studies – one quantitative and one qualitative in nature. Study 1 employed hierarchical generalized linear modeling to simultaneously examine the effects of respondent characteristics and experimentally manipulated item characteristics on the probability of midpoint response option endorsement. Respondent characteristics included self-reported effort expended on the assessments administered and respondent levels of verbal aptitude (SAT verbal scores). Respondents were

randomly assigned different forms of the instrument which varied in item set location (scales administered earlier versus later in the instrument) and midpoint response option label (unlabeled, *neutral*, *undecided*, *neither agree nor disagree*). Experimental manipulation of these variables allowed for a stronger examination of these variables' influence and how they might interact with respondent characteristics (i.e., effort, verbal aptitude) relative to previous studies investigating the issue. Study 2 employed a think-aloud protocol to further examine and understand respondent use and conceptualization of the midpoint response option upon manipulation of midpoint response option label (unlabeled, *neutral*, *undecided*, *neither agree nor disagree*). Four female and four male participants were randomly selected to participate in the think-aloud process using a subset of the items administered in Study 1.

Findings from both studies suggest that the midpoint response option is prone to abuse in practice. Results of Study 1 indicate that respondent characteristics, the experimental manipulation of item characteristics, and their interactions have the potential to significantly affect probability of midpoint response option endorsement. Results of Study 2 reveal that justifications provided by respondents for midpoint response endorsement are mostly construct-irrelevant and differences in conceptualization of the midpoint response option across variations in label appear to be idiosyncratic. These findings have significant implications for the validity of inferences made based upon noncognitive assessment scores and the improvement of assessment practice.

I. Introduction

Chapter Overview

Although response scales for Likert items containing a midpoint response option are used in instruments administered in wide variety of contexts, this chapter begins by describing their use in noncognitive instruments administered for the purposes of higher education assessment. The importance of garnering validity evidence for noncognitive instruments utilized in higher education assessment is then described with particular attention given to how validity might be affected when administering a scale (or items) including a midpoint response option. Because the majority of studies investigating midpoint response selection have been conducted in the survey research context, this chapter contrasts surveys with noncognitive assessments and concludes that the findings from the survey research literature may generalize to noncognitive assessments. The relevant literature from survey research pertaining to response processes and response styles is briefly described. Particular attention is paid to midpoint response style and studies investigating the relationships between item and respondent characteristics that potentially influence exhibition of response patterns consistent with midpoint response style. The chapter concludes with an argument for the current study, which seeks to build upon prior research by examining the role of various item and respondent characteristics on the probability of midpoint response option endorsement.

Noncognitive Assessment in a Higher Education Context

Calls for greater accountability and transparency in higher education from external stake-holders (Spellings, 2006) have been met with an increase in the implementation of assessments to evaluate a multitude of student learning outcomes

(Erwin, 1991; Ewell, 2002). Aggregate results from these assessments are often reported to external audiences for accountability purposes but are also utilized by higher education institutions to inform their curriculum and teaching practice in order to improve student learning and development. Historically, the primary focus of assessments implemented for accountability purposes in higher education has been the assessment of performance on various student learning outcomes associated with general education and major-specific academic programs. Many of these student learning outcomes are strictly knowledge-based and cognitive in nature; thus they focus on the evaluation of student ability in areas such as quantitative and scientific reasoning, written and oral communication, and major-specific content knowledge.

However, institutions recognize that there is an affective component to student learning and development. Hence, many institutions have invested in the evaluation of student development on various noncognitive outcomes as well. In contrast to knowledge-based, cognitive outcomes, noncognitive outcomes typically refer to a variety of affective and attitudinal characteristics such as student levels of tolerance, openness, empathy, and civic responsibility. The fostering and development of such noncognitive characteristics has long been a goal of higher education institutions, but only more recently has there been a push for the assessment of these outcomes. This renewed interest in assessing noncognitive student outcomes has been accompanied by more and more empirical evidence to suggest that these outcomes contribute to student success but are largely distinct from cognitive ability (Kyollen, 2005; Levine-Brown, Bonham, Saxon, & Boylan, 2008; Saxon, Levine-Brown, & Boylan, 2008). Thus institutions are

becoming increasingly interested in devoting time and resources to the evaluation of student development on a variety of affective and attitudinal characteristics.

Noncognitive student outcomes are commonly assessed via self-report measures that are associated with Likert-type response scales associated with qualifying labels. Typically, students are prompted to respond to specified statements using the Likert response scale to indicate their level of agreement or disagreement with each statement. For example, a student may be presented with an item statement that reads “I have a clear set of personal values or moral standards” and corresponds to a 7-point Likert-type response scale labeled, 1 “Strongly Disagree”, 2 “Disagree”, 3 “Slightly Disagree”, 4 “Neutral”, 5 “Slightly Agree”, 6 “Agree”, and 7 “Strongly Agree”. Students then utilize the 7-point response scale to indicate their level of agreement or disagreement with the statement. Likert-type response scales, such as the one described above, inherently suggest that student levels of agreement fall along a continuum. That is, a student selecting response option 1 “Strongly Disagree” would fall at one polar end of the agreement continuum indicating the lowest level of agreement with a given statement, whereas another student selecting response option 7 “Strongly Agree” would fall at the opposite polar end of the agreement continuum indicating the highest level of agreement with the statement. Accordingly, a student selecting response option 4 “Neutral” would fall in the middle of the agreement continuum indicating a middling level of agreement with a given statement whereas a student selecting a response slightly below (3 “Slightly Disagree” or 2 “Disagree”) or slightly above (5 “Slightly Agree” or 6 “Agree”) 4 would fall towards either polar end of the agreement continuum indicating a slightly lower or higher level of agreement with a given statement.

Institutions utilize data collected from these self-report measures with Likert-type response scale options in order to draw inferences regarding student development on a variety of noncognitive outcomes. These findings are then used to inform practice in promoting and addressing these outcomes in the higher education context. Importantly, the quality of the inferences one can make based upon this kind of data depends on the accuracy of student responses and in turn, the accuracy of student responses is based upon the degree to which their response selection accurately reflects their underlying levels of a given construct the instrument purports to measure. That is, a student exhibiting a high level of a noncognitive construct, such as motivation to learn, should respond accordingly by selecting a positive Likert-type response scale option. Correspondingly, a student exhibiting a middling level of motivation to learn should respond accordingly by selecting the midpoint Likert-type response scale option. However, is it possible that a student's response selection may not reflect his or her underlying level of the noncognitive construct of interest? Instead could a student's response selection reflect something else entirely such as the effort put forth on the assessment? And if so, what does this mean for the inferences that institutions would like to make based on student scores on noncognitive assessments?

Response Error and Validity of Respondent Scores

The discrepancy between a respondent's answer selection and his or her most accurate response option selection is often referred to as "response error" (Sudman & Bradburn, 1974). Sources of response error in noncognitive assessment are potentially infinite, meaning that students could provide responses that differ from their true response for any number of reasons. The effects of these multiple sources of response

error can potentially contribute to systematic variance in respondent scores that is construct-irrelevant, or unrelated to the construct of interest that an instrument purports to measure. The presence of construct-irrelevant variance within respondent scores is considered to be a validity issue as it diminishes the strength of the conclusions one can make based upon respondent scores.

The 1999 version of the *Standards for Educational and Psychological Testing* defines validity as, “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9). Thus, if educational researchers and assessment practitioners in higher education wish to make valid inferences regarding student development on noncognitive outcomes, then they must work to garner evidence to support their conclusions. This includes providing evidence to support the inference that student scores on noncognitive assessments accurately reflect levels of the noncognitive constructs these instruments purport to measure. In order to ensure that these kinds of inferences can be made, researchers and assessment practitioners alike must examine any potential sources of systematic construct-irrelevant variance that may reduce the validity of student scores.

Much research has been conducted in the area of survey methodology to examine the effects of certain construct-irrelevant factors that potentially influence respondent scores. More specifically, researchers in the field of survey development and administration have long been interested in examining the effects of response scale characteristics on survey participant responses. Response scale characteristics hypothesized to be sources of construct-irrelevant variance range in specificity anywhere from the number of response scale options offered (see Presser & Schuman, 1980; Weng,

2004) or labeling of response categories (see Weng, 2004) to different color options for lower and upper bounds of a computer-based survey response scale (see Tourangeau, Cooper, & Conrad, 2007). However, there is one long-debated characteristic of the Likert response scale associated with potential sources of construct-irrelevant variance that has been examined extensively within the field of survey research and yielded contradictory results - the midpoint response option.

The midpoint response option. Likert-type response scale midpoint options, sometimes left unlabeled, but often labeled *neutral*, *undecided*, or *neither agree nor disagree*, are commonly used in both survey research and noncognitive assessment. Most survey researchers and assessment practitioners score this response scale selection as the midpoint along a continuum (e.g., assigning a score of 3 for midpoint response on 5-point Likert scale). However, this may not always be accurate, which in turn, could pose a potential threat to the validity of inferences made based upon respondent scores. That is, participant midpoint responses may not reflect participant middling levels of the construct alone, but could be confounded by other factors related to characteristics of both the survey or assessment instrument and/or the respondent's disposition.

Imagine that a student completes a noncognitive assessment which includes an item statement that reads, "I have a definite sense of purpose in life", with a corresponding response scale that ranges from 1 to 7 labeled: 1 "Strongly Disagree", 2 "Disagree", 3 "Slightly Disagree", 4 "Neutral", 5 "Slightly Agree", 6 "Agree", and 7 "Strongly Agree". The student may respond to the item by selecting the midpoint response option – 4 which is labeled "Neutral". This student's response selection would typically be assigned a score of 4, which falls directly in the middle of the 7-point Likert-

type response scale, indicating the student's middling level of sense of purpose in his or her life. However, this may not actually be the case. That is, this student's midpoint response selection could be indicative of any number of factors unrelated to a middling level of the construct of interest such as item clarity, item positioning, response option labeling, or student levels of motivation to complete the instrument. Thus, the purpose of the current study is to better understand what construct-irrelevant factors influence student selection of the midpoint response in a higher education noncognitive outcome assessment context by building upon prior research conducted within the field of survey methodology. Moreover, I hope to expand upon this research by further examining student conceptualization of the midpoint response option through qualitative analysis.

Survey Instruments versus Noncognitive Assessments

As mentioned previously, most of the research conducted in this area has not focused on examining the construct-irrelevant factors that contribute to response error in the noncognitive assessment context. Instead, much of this research has been conducted in the field of survey methodology. Although surveys are similar to noncognitive assessments in that they typically utilize a Likert-type response scale, they often differ in their purpose (i.e., the material they purport to measure) and the inferences one makes based upon respondent scores.

Survey items commonly refer to respondent attitudes towards or opinions on a particular issue or the frequency of respondent behaviors in order to provide details of public beliefs, attitudes, and practices. Conversely, noncognitive assessment items typically refer to respondent traits or characteristics in order to provide an estimate of respondent underlying levels of the construct of interest. The major distinction between

the two types of items is that a response to a single survey item in and of itself is often of interest, whereas a response to a single noncognitive item is not often considered in isolation. Noncognitive items are typically administered along with a collection of items measuring the same construct. Thus it is the level of the construct established by response selections across multiple items measuring the same construct, not response to any one item that is of interest.

Despite the distinctions amongst the purposes of and inferences made based upon responses to survey versus noncognitive assessment items, response processes may be quite similar across both. That is, although survey and noncognitive items typically differ in purpose, item content, and number of items administered measuring a given construct, there is little reason to suspect that response processes would vary substantially across these two item types. Accordingly, research conducted in the field of survey methodology may be useful for informing research that has yet to be conducted in the field of noncognitive assessment.

Response processes. Tourangeau, Rips, and Rasinski (2000) propose a model of the response process in the context of survey administration that seems as though it would also apply in noncognitive assessment. According to Tourangeau and colleagues, the typical response process for a given survey item consists of four components: respondent “comprehension of the item, retrieval of relevant information, use of that information to make required judgments, and selection and reporting of an answer” (Tourangeau et al., 2000, p. 7). The multifaceted nature of this model implies that error may be introduced into respondent scores at multiple points due to a myriad of different factors.

Survey methodologists have conducted extensive research that attempts to examine both the causes and effects of error in survey responses. This research has been conducted with the intention of informing survey development and administration to avoid and/or control for bias in respondents' scores. Findings suggest that there is a multitude of item and respondent characteristics that can potentially bias respondent scores. These findings appear to be logically consistent with the response process model that Tourangeau et al. (2000) proposed. More specifically, both item and respondent characteristics could potentially influence any step of their proposed cognitive response process, introducing error in respondent scores. For example, it is easy to imagine that item comprehension may be influenced by item characteristics such as item statement wording, length, and complexity and/or respondent characteristics such as level of reading comprehension, the extent of vocabulary, or language preference. If comprehension of the item statement is indeed influenced by any of these irrelevant characteristics, then response to the item may be inaccurate which calls into question the validity of the inferences that one can make based upon respondent scores.

Response styles. One potential source of irrelevant variance in respondent scores that has been studied extensively are response styles, otherwise referred to as response sets, or response biases within the survey research literature (Baumgartner & Steenkamp, 2001; Cronbach, 1946, Cronbach, 1950; Harzing, 2006; Kieruj & Moors, 2010). Response styles refer to the tendency of participants to respond in a systematic manner across items independent of survey content (Baumgartner & Steenkamp, 2001; Cronbach, 1946, Cronbach, 1950; Harzing, 2006; Kieruj & Moors, 2010). There are a number of different response styles that have been examined within the survey literature including:

acquiescence response style (ARS), disacquiescence response style (DARS), extreme response style (ERS), and midpoint response style (MRS). Each response style is unique in that it corresponds to a different systematic response pattern, yet similar in that each response pattern remains consistent across items despite differences in content.

According to Cronbach (1946), the presence of construct-irrelevant variance in respondent scores introduced by response styles may affect the psychometric properties of the instrument, such as reliability and validity, in inconsistent ways. That is, estimates of reliability (e.g., Cronbach's Alpha, test-re-test, alternate forms) and validity (e.g., correlation estimates with external criteria) may be artificially biased either upwards or downwards depending on the situation (Cronbach, 1946). However, response styles pose an even greater threat to validity in that, "they permit persons with equal knowledge, identical attitudes, or equal amounts of a personality trait to receive different scores" (Cronbach, 1946, p. 491). This ultimately affects the quality of the inferences one can make based upon the data as the researcher cannot disentangle sources of true variation due to underlying levels of ability on a construct from construct-irrelevant variance due to response styles.

Item and respondent characteristics. Research suggests that there are several different factors related to characteristics of the measure, as well as participant characteristics or dispositions that may influence response styles (Baumgartner & Steenkamp, 2001; Hamilton, 1968; Harzing, 2006; Kieruj & Moors, 2010; Velez & Ashworth, 2007). Furthermore, there is evidence to suggest that these factors relate differentially to various classifications of response styles such as extreme response style

(ERS) and midpoint response style (MRS) (Hamilton, 1968; Harzing, 2006; Kieruj & Moors, 2010; Krosnick, 1999).

Given the findings of this research, there has been some initial examination of item and respondent characteristics that potentially influence midpoint response option endorsement in a noncognitive higher education assessment context. To this end, Marsh and Pastor (2011, 2012) investigated the effects of a series of item characteristic variables (i.e., item position, item reverse scoring, item midpoint response option label) and respondent characteristic variables (i.e., respondent verbal ability, respondent self-reported level of effort put forth on the instrument, and self-reported level of their perceived importance of the instrument) on likelihood of midpoint response option endorsement. Results of the exploratory Marsh and Pastor studies (elaborated upon in Chapter 2) suggest that some item and respondent characteristics have the potential to introduce construct-irrelevant variance within respondent scores by effecting endorsement of the midpoint response option; however findings across the two studies were not consistent.

Purpose of the Current Study

To date, most survey research studies have examined the relationship amongst only one or a few possible measure-specific and/or respondent dispositional characteristics and midpoint response endorsement, possibly failing to investigate the nature of the relationship between any number of these variables and midpoint response after controlling for other variables within the model. Moreover, much of the research conducted in this field does not focus on the influence of these variables in a noncognitive assessment context. Although response processes may not differ

substantially across survey and noncognitive items, we cannot be certain that survey research findings will generalize across contexts. Furthermore, we do not understand the specifics of when and how measure-specific and respondent dispositional characteristics potentially influence midpoint response option endorsement.

Thus, the purpose of the current study was to expand upon this body of research by simultaneously examining the effects of a number of construct-irrelevant item and respondent characteristics that potentially affect likelihood of midpoint response selection and to better understand student conceptualization of the midpoint response option on noncognitive assessments utilized in a higher education context. Upon establishing a better understanding of respondent use of the midpoint response option and the factors that influence its potential misuse in practice, researchers may be able to better control for bias in respondent scores. A mixed-method experimental design which requires collection and analysis of both quantitative and qualitative data was employed in order to fully address the research questions associated with this purpose statement.

This mixed-methods approach was executed by way of two studies. Study 1, which required the collection and analysis of quantitative data, examined various item and respondent characteristics that may influence respondent selection of the midpoint response option. Study 1 participants were randomly assigned to one of eight forms of a noncognitive assessment instrument associated with different item characteristic manipulations including item set location (earlier versus later within the assessment) and item midpoint response option label (no label, *neutral*, *undecided*, *neither agree nor disagree*). Study 1 respondent characteristics included respondent self-reported levels of effort expended on assessments and verbal aptitude (SAT verbal scores). Quantitative

data utilized in Study 1 was collected from a university-wide Assessment Day testing session and analyzed using hierarchical generalized linear modeling. This approach allowed the researcher to address the multifaceted nature of the research questions associated with Study 1. More specifically, hierarchical generalized linear modeling allows for the effects of both item and respondent characteristics on participant likelihood of midpoint response endorsement to be estimated simultaneously. Selection and inclusion of item and respondent characteristics within the specified model was informed by findings from previous studies conducted in the area of survey research and addresses gaps within the research literature.

Study 2, which required the collection and analysis of qualitative data, further examined respondent use and conceptualization of the midpoint response upon manipulation of the midpoint response option label. More specifically, Study 2 required eight participants to engage in a think-aloud procedure using a subset of the noncognitive assessment items administered in Study 1. Think-aloud item sub-groupings were associated with a different experimentally manipulated midpoint response option label (either, no label, *neutral*, *undecided*, *neither agree nor disagree*). Item groupings and their associated midpoint response option label were consistent across the eight participants, but order of item grouping administration differed such that each participant received a different form. During this think-aloud procedure, the researcher prompted participants to articulate response processes when responding to each item administered. Following the think-aloud, the researcher asked specific questions regarding the manipulation of the midpoint response option label and its influence on participant midpoint endorsement. Participant responses were audio-recorded by the researcher and

then analyzed and coded for major themes that emerged regarding general use of and conceptualization of the midpoint response option across different manipulations of the label. Qualitative data resulting from the think-aloud procedure was used to further expand upon and supplement findings from quantitative data collected in Study 1.

II. Review of the Literature

Chapter Overview

The current section will begin with general discussion of midpoint response option use in assessment and measurement practice. Then the delineation will be explained between proper midpoint response option use versus abuse in a noncognitive assessment context. Discussion of research examining midpoint response option abuse in assessment and measurement practice will follow. Next will be a review of response style research focusing on the effects of various item and respondent characteristics on midpoint response style, followed by a review of exploratory research that builds upon these findings examining the effects of item and respondent characteristics on probability of midpoint response endorsement. Finally, this section will close with a description of the research questions addressed within this dissertation.

Proper Use versus Abuse of the Midpoint Response Option in Practice

The utility and necessity of a midpoint response option has long been discussed in survey research literature. Consistent with practice in the field of educational measurement, survey researchers emphasize that decisions regarding the specifics of survey development and administration, such as the number and nature of response options offered, should be dependent upon the kind of inferences the researcher wishes to make based on participant responses. Thus although there seems to be no definitive conclusion on whether a midpoint response option should be offered or excluded on a Likert-type response scale within the survey research literature, it does follow that the decision should be dictated by the kind of inferences the researcher wishes to make based upon participants' responses to item statements.

In accordance with this perspective, Payne (1951) suggests that the decision to offer or exclude a midpoint response option in survey administration should be dictated by the kind of information or inferences the researcher wishes to make based upon participant response. More specifically, Payne (1951) advises that researchers should consider offering the midpoint response option when they wish to gain more definitive information regarding participant responses on a certain issue, but consider excluding the midpoint response option when they wish to obtain information regarding which direction respondents tend to lean on a particular issue. Presser and Schuman (1980) build upon Payne's perspective by describing three common assumptions researchers often make when deciding whether to include or exclude the midpoint response option on a response scale:

Three hypotheses are implied by the way the middle position is handled in the wording of questions. First, when survey investigators decide against offering an explicit middle alternative, they are usually assuming that the middle category consists largely of responses from those who lean toward one or the other polar alternatives, though perhaps with little intensity. Thus it is legitimate to press respondents to choose one of these alternatives, rather than allowing them to take refuge in a middle position. Second, some investigators omit the middle alternative in the belief that it tends to attract people who, having no opinion on the issue, find it easier to choose a seemingly noncommittal position than to say "don't know." Third, investigators who do offer a middle alternative are probably assuming that respondents who opt for it really do favor the middle position, and

if forced to choose a polar alternative will contribute some form of random or systematic error to the distribution. (Presser & Schuman, 1980, p. 71)

Presser and Schuman's (1980) description of these assumptions reflects a distinction between two kinds of uses of the midpoint that will be used throughout this dissertation to characterize respondent behavior. The first kind of use is termed "proper use" of the midpoint. When respondents use the midpoint properly, they select the midpoint response because they truly have "equal feelings of agreement and disagreement" (Harter, 1997, p. 17). Both Payne (1951) and Presser and Schuman (1980) imply that researchers deciding to offer the midpoint response option assume that respondents are using the midpoint response option appropriately or *properly*. They suggest that these researchers opt to offer a midpoint response option because they believe that a respondent's endorsement of this option accurately reflects preference for the middle position on a given issue. If respondents intend to use the midpoint properly, but it is not provided as an option by the researcher, less "definite information about a respondent" could be obtained and, even worse, "some form of random or systematic error" could be introduced into scores by forcing them to choose a polar alternative (Presser & Schuman, 1980, p. 71).

When a researcher decides to offer a midpoint response option, it is with the hope that it will be properly used. However, as alluded to by Presser and Schuman, midpoints are often excluded because it is assumed that the respondent will use the midpoint response inappropriately, thus introducing error in responses. This second kind of use of the midpoint is termed "abuse" of the midpoint. Fear of midpoint "abuse" is most often the reason many researchers opt to exclude the midpoint response option. They fear that respondents selecting the midpoint may actually lean towards the polar options or that

respondents with no opinion on the matter will be more inclined to endorse the midpoint rather than another more appropriate option (such as an additional, *no opinion* response option).

Midpoint response option abuse, as defined in this dissertation, is specifically characterized by the error introduced within respondent scores when midpoint response endorsement fails to accurately correspond to its scoring. That is, based on typical Likert-type response scale scoring (e.g., assigning a score of 4 to a midpoint response option on a Likert-type scale that ranges from 1-7), midpoint endorsement is intended to reflect equal levels of respondent agreement and disagreement to an item statement or middling levels of a given construct an item purports to measure. Error introduced within respondent scores due to midpoint response abuse results when midpoint response option endorsement fails to reflect this, rendering typical scoring inappropriate or invalid.

According to Harter (1997), survey research suggests that respondents commonly associate the midpoint response option, sometimes referred to as a nondirectional response in the survey research literature, with one of three meanings:

- (1) ambivalence – “expressed when a person endorses both positive and negative attitudinal positions” (Breckler, 1994, p. 350)
- (2) uncertainty (not knowing) – “you are just not able to identify your feelings” (Klopper et al., 1980, p. 98); and
- (3) indifference (neutrality) – “expressed lack of concern or apathy with respect to the issue” (Edwards, 1946, p. 162), “you don’t care one way or the other”, (Klopper et al., 1980, p. 98). (Harter, 1997, p. 17)

Harter's summary of these various conceptualizations associated with the midpoint response option illustrates one instance when midpoint endorsement potentially fails to accurately reflect its typical scoring. That is, according to the definitions of proper midpoint response option use versus abuse provided above, midpoint response endorsement reflecting ambivalence or "when a person endorses both positive and negative attitudinal positions", would be considered proper midpoint response option use as it most accurately reflects a respondent's middling stance on a given issue thus corresponding to its typical scoring. Conversely, midpoint response endorsement reflecting either respondent uncertainty or indifference would be considered abuse as their associated meanings provided by Klopfer et al. (1980) and Edwards (1946) indicate that they reflect respondent ignorance or apathy.

An additional instance of midpoint response option abuse can result from a response behavior that Krosnick (1999) refers to as *satisficing*, or the act of selecting a response option that requires less cognitive effort on behalf of the respondent. Krosnick (1999) subscribes to the same four-step cognitive response process that Tourangeau et al. (2000) detail and emphasizes the cognitive complexity of this process, stating that the optimal survey item response requires respondents to "interpret the question and deduce its content...search their memories for relevant information and then integrate that information into a single judgment...and translate the judgment into a response by selecting one of the alternatives offered" (p. 546). According to Krosnick, respondents are *optimizing* when they exert the appropriate amount of effort required to perform each of these processes and arrive at the most accurate response. Alternatively, participants are said to be *satisficing* when they only partially engage in this response process or devote

less effort to each step of the process resulting in a response selection that the respondent may deem to be satisfactory, but is ultimately less accurate.

Error introduced within respondent scores as a result of satisficing is not unlike error introduced within respondent scores due to differences in respondent conceptualization of the midpoint response option and its typical scoring as previously described. Although the mechanism influencing midpoint response option abuse differs (e.g., discrepancies in meaning associated with the midpoint response option and its scoring versus satisficing), the result is similar. In both circumstances, midpoint response endorsement does not accurately reflect a middling stance on a given issue or a corresponding level of a noncognitive construct, introducing construct-irrelevant variance within respondent scores. In turn, this construct-irrelevant variance results in error that threatens the validity of the inferences researchers make based on their data.

Midpoint Response Option Abuse in Assessment and Measurement Practice

In order to avoid the introduction of error within respondent scores due to either scenario, it is imperative that researchers attempt to better understand the factors and/or circumstances that influence proper midpoint response use versus abuse. This in turn, will help inform best practice in deciding whether to offer or exclude a midpoint response option and allow researchers to control for potential confounds if and when they do decide to offer the midpoint. But before researchers can investigate the specifics of various sources of construct-irrelevant variance that potentially threaten the inferences they make, there must be empirical evidence to support that the midpoint is abused.

DeMars and Erwin (2004) employed item response theory (IRT) methods in order to examine whether college student selection of a midpoint response option labeled

neutral or unsure corresponded to middling self-reported levels of identity development. The nominal item response model utilized by DeMars and Erwin, “uses a function that describes the probability at different levels of development of choosing each response” allowing them to determine how students across varying levels of identity development use each response option (p. 89). Through this application of the nominal response model, DeMars and Erwin determined that the midpoint response option labeled *neutral or unsure* was used differentially across identity development subscales and items. In other words, the highest probability of midpoint response option endorsement was not consistently associated with students exhibiting middling levels of self-reported identity development. Instead, the highest probability of midpoint response option endorsement was associated with differing levels of self-reported identity development across subscales and items, suggesting that the midpoint does not consistently reflect middling levels of student identity development.

Like DeMars and Erwin (2004), Harter (1997) also used IRT methods to examine whether offering a sixth *don't know* response option to a 5-point Likert-type response scale would improve functioning of the midpoint response option. Harter hoped to minimize error and deter midpoint response option abuse by offering an additional *don't know* option for indifferent and uncertain respondents potentially drawn to the midpoint. His findings generally supported his hypothesis as the psychometric properties of the *undecided* midpoint response option were improved upon when an additional *don't know* response option was present. More specifically, “the middle response alternative may function more like a theoretically true middle response”, given the levels of theta

consistently associated with endorsement of the *undecided* when an additional *don't know* option is offered (Harter, 1997, p. 125).

Midpoint response option label and midpoint abuse. As evidenced by the previous two studies, midpoint response options can and often do differ in the labels assigned to them. In the DeMars and Erwin (2004) study, the midpoint response option was labeled *neutral or unsure* versus *undecided* in the Harter (1997) study. Midpoint response option labeling conventions utilized in survey research commonly include variations such as, *neutral, unsure, undecided, neither agree nor disagree, no opinion, don't know*, and sometimes no label. Although differences in labeling conventions across instruments are prevalent in practice, there is often little or no justification provided by researchers to justify their selection. However, according to DeMars and Erwin (2004, 2005), there are inherent differences in meaning associated with each labeling convention that have the potential to influence conceptualization and scoring of the midpoint response option:

Taken literally, *neutral* would seem to indicate a middle level of endorsement, while *no opinion* or *unsure* would seem to indicate a lack of an opinion or a lack of interest in the topic...The neutral response category falls under the broader classification of middle response options. Another example of a middle response option would be *about right* in a question where the options were *too much, not enough, or about right*. These options are not necessarily placed in the middle of the list of options, but they are interpreted to mean, and scored as, middle-level opinions. *Unsure, no opinion, cannot decide, and don't know* would seem to be a different class of options; they do not necessarily indicate a middle position but

may be used when the respondent has no opinion or lacks enough information to form an opinion. (DeMars & Erwin, 2005, p. 2)

If the label of the midpoint response option does in fact impact its conceptualization and scoring as DeMars and Erwin suggest, it would serve as an additional source of construct-irrelevant variance, possibly biasing respondent scores. For instance, imagine a survey in which midpoint response options are labeled *neutral* for some items and *undecided* for others. The same respondent may interpret a midpoint response option labeled *neutral* differently than one labeled *undecided*. In a noncognitive assessment context, this respondent may interpret a midpoint response option labeled *neutral* to reflect a middling level of the construct of interest whereas he or she may interpret a midpoint response option labeled *undecided* to reflect non-response. Moreover, two different respondents may interpret the same midpoint response option label differently. Perhaps one respondent interprets an item with a midpoint response option label of *neutral* to reflect his or her middling level of the construct of interest whereas another interprets this midpoint label for the same item to indicate uncertainty in level of the construct of interest. These discrepancies pose a threat to the validity of respondent scores when the midpoint response option is scored to reflect a middling level of a given construct. This is due to the fact that researchers can no longer definitively conclude that midpoint response endorsement represents a middling level of the construct of interest given potential variations in respondent interpretation across labels.

Although there is reason to speculate that subtle nuances in the meaning associated with different response option labels may alter interpretation and endorsement of the midpoint response option, there has been little empirical study to support or refute

this claim. One study conducted by Bishop, Hippler, Schwarz, and Strack (1988) noted a difference in the proportion of respondents who endorsed the midpoint response option versus an additional, *no opinion* response option across two different forms of a survey (one that offered the additional *no opinion* option versus another that offered only the midpoint response option) suggesting that respondents discriminate between a midpoint response option and an additional *no opinion* response option. Findings from studies such as Bishop et al. (1988) and Harter (1997) appear to suggest that respondents can and do discriminate between response options with different labels and similar meanings such as *undecided* versus *don't know* or *no opinion*. While these studies are helpful in determining whether response selection is influenced by response option label, they do not provide definitive evidence as to if or how midpoint response option endorsement is influenced by systematic variation in traditional labels.

In attempts to address this gap in the research literature and build upon the findings of their previous study examining the appropriate scoring of a single *neutral or unsure* midpoint response option, DeMars and Erwin (2005) sought to determine whether respondents selecting a midpoint response option labeled *neutral* versus *unsure* differed in their levels of identity development. Their results were similar to their 2004 study in that levels of identity development associated with midpoint response endorsement were not necessarily associated with middling levels of identity development. However, interestingly, their results also indicated that respondents selecting either of these midpoint response options did not appear to differ significantly in their underlying levels of identity development, suggesting that respondents use these two labels interchangeably despite their different labels.

Overall, findings from the Harter (1997), Bishop et al. (1988), and DeMars and Erwin (2005) studies are rather contradictory. Findings from the former suggest that respondents can and will discern amongst midpoint response options even across labels with seemingly similar meanings, but results from the latter indicate that this does not appear to be the case (at least with regard to midpoint response options labeled either *neutral* or *unsure*). But what about other potential variations in midpoint response option label (e.g., *neutral*, *undecided*, *neither agree nor disagree*, or no label) - would these response options be used interchangeably as well? Given the discrepant nature of these findings and paucity of research conducted in this domain these kinds of questions warrant further investigation. More specifically, it is important to fully investigate if and how different midpoint response option labels influence respondent conceptualization and endorsement of this response option in order to determine if and how this response scale characteristic contributes to midpoint response option abuse.

In addition to the IRT methods previously discussed, there have been numerous studies conducted utilizing Classical Test Theory (CTT) approaches in order to determine whether the midpoint response option is abused in practice. This research most notably focuses on examining the psychometric properties (i.e., validity and reliability) of respondent scores when such an option is included versus excluded within a survey response scale. Much like the studies examining the relationship between midpoint response option label and midpoint abuse, findings from these studies have historically been somewhat contradictory. That is, there has been some evidence to suggest that increasing the number of Likert scale response options by offering a midpoint bolsters the reliability and validity of respondent scores by allowing them to more accurately

represent their stance along a response continuum (Nunnally, 1967; O’Muircheartaigh, Krosnick & Helic, 2000; Rugg & Cantril, 1944). In contrast, there has been counter evidence to suggest that offering the midpoint response option threatens score reliability and validity, introducing error within scores by attracting respondents likely to abuse this option. As mentioned previously, most researchers opting to exclude the midpoint response option have reason to believe that this will be the case. Those researchers who opt to exclude the midpoint response option may cite numerous studies that suggest that the act of merely offering a midpoint response option significantly increases respondent selection of this option (Kalton, Roberts & Holt, 1980; Klopfer & Madden, 1980; Presser & Schuman, 1980; Si & Cullen, 1998).

Midpoint Response Style (MRS) Research

While the studies previously described provide evidence to suggest that the midpoint response option may be subject to abuse, there is still much research to be done investigating the specifics of the conditions that potentially influence midpoint response option abuse. In other words, given the contradictory nature of the findings of the studies described above, it appears that midpoint response option abuse may be a potential threat to the validity of respondent scores in some conditions, but not in others. Thus it is imperative to examine under what conditions (both measurement- and respondent-specific) are respondents likely to abuse the midpoint. Knowledge of these potentially biasing factors would not only provide additional evidence to support or refute midpoint response option abuse but inform researchers and practitioners alike of the conditions that foster midpoint response abuse in practice. Upon a better understanding of the nature of these factors and their effects, changes may be implemented within survey and

assessment methodology in order to diminish or potentially eliminate error due to midpoint response option abuse.

Much of the research examining factors such as these has been conducted in an area of survey research focusing on the specific conditions that contribute to what may be considered a more pervasive form of midpoint response option abuse called midpoint response styles (MRS). MRS refers to a specific kind of construct-irrelevant response behavior that manifests itself in a “disproportionate use of the middle response category” (Kieruj & Moors, 2010, p. 321). Recall that within this dissertation, midpoint response option abuse has been characterized by error introduced within respondent scores when a respondent selects the midpoint response option in order to convey anything other than ambivalence or a middling level of the construct of interest for any given item. Accordingly, MRS would be considered a more extreme case of midpoint response abuse in that it consistently endures across items and measures. Despite this distinction and the fact that these MRS studies have been primarily conducted in a survey administration context, these findings may inform research that has yet to be conducted in a noncognitive assessment context by foreshadowing the effects of different characteristics that potentially contribute to midpoint response option abuse. Thus the following sections review these studies in an attempt to inform research yet to be conducted in a noncognitive assessment context.

Response style research, and MRS research specifically, has primarily focused on examining the role(s) of different item and respondent characteristics that relate to this kind of response behavior. Yet before these effects can be examined, researchers must take care to establish that consistent response patterns characteristic of response styles are

indeed a reflection of construct-irrelevant variance. With regard to MRS, this means that researchers must take measures to design their studies in such a way as to ensure that consistent midpoint response option endorsement is due only to a respondent's middling stance, attitude, or middling level of a construct of interest. In order to address this concern and avoid confounding what Baumgartner and Steenkamp refer to as "stylistic variance with substantive variance", response style researchers have typically administered a purposeful selection of items measuring different content (Baumgartner & Steenkamp, 2001, p. 144). The justification for this approach is that it largely diminishes the probability that a respondent will respond consistently across an assortment of items that are unrelated in content. That is, a respondent is likely to respond similarly across a series of items inquiring about attitudes towards a single issue (e.g., attitudes towards capital punishment) or series of similar issues (e.g., attitudes towards different environmental policies), but is rather unlikely to respond similarly across items that are heterogeneous (e.g., some items inquiring about attitudes towards health care reform and some items inquiring about attitudes towards household pets). Thus the majority of the following studies have taken this or similar approaches to control for this potential confound and examine the role of item and respondent characteristics in relation to MRS.

Item Characteristics and MRS

Hoping to build upon previous research indicating a link between item statement ambiguity and participant *no opinion* response option selection (Coombs & Coombs, 1976), Velez and Ashworth (2007) designed a study to examine participant midpoint response selection across items of differing levels of readability (i.e., reading grade level and perceived clarity of item statement) when there was no additional *no opinion* option

offered to respondents. Velez and Ashworth (2007) analyzed a series of item readability statistics including the Flesch-Kincaid Grade Level Index for each survey item statement and expert ratings of item perceived clarity in order to fully evaluate item readability and its relationship to midpoint response option endorsement. They discovered that as item readability decreased and item statements became more complex and consequently more difficult to read, perceived clarity decreased. Additionally, participant endorsement of the midpoint response increased as readability and perceived item clarity decreased, providing evidence to suggest that, “some of the data may represent respondent confusion based on lack of understanding of the item or the respondent’s unwillingness to expend a great deal of cognitive effort to interpret the survey question, resulting in systematic error in one’s data” (Velez & Ashworth, 2007, p. 72).

Kieruj and Moors (2010) conducted a study to identify survey respondents exhibiting either extreme response styles (ERS) or MRS across manipulations of one method factor that they believed to potentially affect the emergence of both – the number of response option categories offered for attitudinal survey items. With regard to MRS, they hypothesized that this kind of response behavior would be more likely to be observed when an odd number of response options was offered compared to an even number of response options (i.e., when a definitive midpoint response option was offered versus when it was not). Interestingly, their findings did not support their original hypotheses associated with MRS. Instead they found that a class of respondents exhibiting MRS “emerged in treatments in which a relatively large number of 9 or 10 answering categories were offered, whereas no MRS emerged if a relatively small number of answering categories were offered” (Kieruj & Moors, 2010, p. 337). This

finding ultimately suggests that respondents were not more likely to exhibit MRS (which according to the conclusions the authors draw from their results can also occur when an even number of response options are provided) due simply to the presence of the midpoint response option among the response categories, but that as item response scale options increase, so too does endorsement of the midpoint response. This trend could be the result of an increase in task difficulty that coincides with an increase in response options, causing respondents to find it increasingly difficult to provide an accurate response. However, it should be noted that this trend did not emerge when the number of response options was increased to 11 (Kieruj & Moors, 2010). Thus, although Kieruj and Moors' (2010) findings provide some evidence to suggest that midpoint response endorsement increases with an increase in response options, the results were inconclusive.

Overall, the findings of these studies provide some evidence to suggest a link between increase in participant midpoint response option endorsement and MRS, specifically to item characteristics such as item readability, perceived clarity, and response scale length. These item characteristics have the potential to increase respondent cognitive load and hinder respondent item statement comprehension. Researchers have long posited that the presence of such item characteristics affects participant response selection and may specifically increase the endorsement of the midpoint response (Bogart, 1967). According to Krosnick, the higher the cognitive load required of the respondent or task difficulty, the greater the chance that they will engage in satisficing. The relationship amongst these item characteristics and increased respondent endorsement of the midpoint may provide further evidence to support Krosnick's claim.

That is, these item characteristics potentially complicate the response process for some respondents and may result in satisficing behavior that manifests itself in the form of midpoint response selection.

In addition to the item characteristics discussed above, there are a number of instrument-specific item characteristics hypothesized to influence response styles whose effects have yet to be examined. For example, some researchers suggest that instrument length and/or the duration of an assessment or testing session may influence respondent selection of the midpoint response option (Krosnick, 1991; Velez & Ashworth, 2007). According to these researchers, instrument length potentially contributes to respondent fatigue that could in turn manifest itself as response styles or satisficing. That is, the longer the instrument, the more likely respondents are to become fatigued, the less effort they are expected to expend on item response, and the more likely they are to select what they may consider to be the easiest or most noncommittal response – the midpoint.

Moreover, there is some research to suggest that item reverse scoring, which systematically alters the wording of items (either positively or negatively), may influence midpoint response selection. Baumgartner and Steenkamp (2001) examined the potential moderating effects of scale balance (“proportion of positively and negatively keyed items” on a scale) on response styles (p. 154). Their findings suggest that scale balance may actually result in a decrease in bias that results from participant response styles. However, as discussed previously, there is alternate research that suggests an increase in midpoint response option selection when an item statement is unclear or confusing (Bogart, 1967). Items that are reverse scored are typically worded in such a manner that endorsement of a higher Likert-scale response option reflects respondent lower levels of a

construct or vice versa. As a result, item statement wording may be more complex, possibly affecting participant midpoint response endorsement.

Finally, as aforementioned, there is also evidence to suggest that midpoint response option labels themselves influence the probability of MRS. Consistent with some of the research discussed previously, there is little direct evidence to suggest that this is the case. However, Harzing (2006) found that different item response scale anchors (“Strongly Disagree” to “Strongly Agree” versus “Of little or no importance” to “Extremely important”) did appear to influence respondent “acquiescence bias” (p. 261), ultimately suggesting that different midpoint option labels may have a similar effect on MRS.

Respondent Characteristics and MRS

Alternate response style research has focused on examining the relationship amongst respondent likelihood to engage in response styles and respondent dispositional characteristics. This research suggests that there are many different dispositional characteristics that can possibly influence participant response in a way that Krosnick claims would result in greater satisficing including: respondent levels of effort (Krosnick, 1999; Velez & Ashworth, 2007), perceived importance of instrument subject matter (Krosnick, 1999; Velez & Ashworth, 2007), cognitive ability (Krosnick, 1999), English language fluency (Harzing, 2006), and various other personality variables (Hamilton, 1968; Harzing, 2006; Krosnick, 1999). According to Krosnick, when respondents wish to expend less effort responding to survey items, they “might offer ‘safe’ answers, such as the neutral point of a rating scale, endorsement of the status quo, or saying ‘don’t know’ so as to avoid expending the effort necessary to consider and possibly take more risky

stands” (Krosnick, 1999, p.548). This type of “safe” or noncommittal response would allow these respondents to provide an answer to a survey item but largely avoid cognitive engagement. Respondents’ levels of perceived importance of survey item content have also been hypothesized to affect their likelihood to satisfice or endorse the midpoint response option (Krosnick, 1999; Velez & Ashworth, 2007). The less important the item content is to respondents, the less likely they are to provide an accurate response either due to a lack of interest in the content or ignorance on the issue. For example, 15 year-old respondents may not perceive a survey item asking about the importance of retirement benefits to be important; thus they may endorse the midpoint response in order to provide a response but avoid expending effort.

Other respondent characteristics such as cognitive ability and English language fluency may also influence endorsement of the midpoint response option. As mentioned previously, Krosnick (1999) suggests that the lower respondents’ cognitive abilities, the more likely they are to engage in satisficing in response to survey items. The belief is that the lower respondents’ cognitive ability, the more likely they are to be confused by an item statement or unable to put forth the cognitive effort required to provide an accurate response. In addition, research suggests that respondent English language fluency also contributes to the cognitive complexity of survey item response, potentially resulting in an increase in midpoint response option endorsement. For example, Harzing (2006) conducted a cross-cultural study examining differences in response styles. Interestingly, she found that as non-native English speakers’ English-language competency decreased, endorsement of the midpoint response on English-language versions of surveys

increased, providing further evidence to support the fact that these respondents are attempting to satisfice by endorsing the midpoint response option.

Both Item and Respondent Characteristics and MRS

Based on the research described above, there is substantial evidence to suggest that both item and respondent characteristics may influence respondent midpoint response endorsement. However, there is a paucity of research examining the potential for item and respondent characteristics to simultaneously influence midpoint response selection. Item characteristics that could potentially influence midpoint responses, independent of levels of the construct the instrument is intended to measure, are often deemed method effects (Kieruj & Moors, 2010). Kieruj and Moors (2010) establish a clear distinction between method effects and response styles, stating, “method effects are attributable to the method itself, whereas response styles appear to be person related features” (p. 322). However, they do acknowledge that although method effects and response styles are “conceptually distinct”, both have the potential to simultaneously influence participant response selection (Kieruj & Moors, 2010, p. 322). Despite the acknowledgement that both item and respondent characteristics potentially play a role in MRS specifically, no research to date has modeled the effects of both types of variables simultaneously. Moreover, the influence of these characteristics on midpoint response option endorsement has yet to be examined in a noncognitive assessment context.

Two preliminary, exploratory analyses were conducted by the researcher and a colleague in order to address this gap in the literature (see Marsh & Pastor, 2011, 2012). These studies specifically sought to predict respondent endorsement of the midpoint response option on items purposefully selected from a series of noncognitive assessments

administered during a university-wide assessment testing administration. In both studies, hierarchical generalized linear modeling was utilized to examine the effects of both item and respondent characteristics simultaneously – building upon previous research conducted in survey methodology that has focused on the relationship amongst only one or a few possible measure-specific characteristics or participant dispositional variables and response styles. Moreover, these studies sought to further this line of research by examining the effects of a wider range of variables on the probability of midpoint response endorsement and by considering the effects of both item and respondent characteristics simultaneously in a noncognitive assessment context.

The data analyzed for both of the Marsh and Pastor (2011, 2012) studies were collected during the same mandatory university-wide assessment testing session held at a mid-sized southeastern institution. Instruments administered and data collected as part of these assessment administrations are utilized to evaluate undergraduate general education and student affairs programming. As such, students are required to participate in these assessment sessions and complete a series of randomly assigned cognitive and noncognitive assessments twice during their undergraduate career, once as incoming freshmen, the Friday before beginning classes and again during their spring semester as sophomore/junior students having acquired between 45-70 course credits. Data from two different subsamples of sophomore/junior students completing an assortment of noncognitive assessments during the same spring 2011 assessment session were analyzed for both studies. In addition, Marsh and Pastor took a similar approach to response style researchers in attempting to control for potential confounds due to construct-relevant responding (i.e., consistent midpoint responding due to middling level of the construct of

interest). More specifically, Marsh and Pastor tried to control for this by including and analyzing responses to a subset of purposefully selected noncognitive assessment items determined to be most distinct from one another in content.

Informed by findings of previous studies conducted in the area of survey research and specific gaps in the literature, Marsh and Pastor (2011) selected a series of item and respondent characteristics they believed to influence the likelihood of midpoint response endorsement on 30 purposefully selected noncognitive assessment items. Item characteristic variables included as predictors of midpoint response endorsement in the first study were: item position (e.g., ordinal position of item on instrument), item reverse scoring (e.g., whether or not an item was reverse scored), item midpoint response option label (e.g., whether the item midpoint response option had a label – *neutral/neither agreement nor disagreement*, or had no label). Person level variables included as predictors of midpoint response endorsement were: respondent verbal ability (e.g., SAT verbal score), respondent self-reported level of effort put forth on the instrument, and respondent self-reported level of perceived importance of the instrument. Although these analyses were ultimately exploratory in nature, Marsh and Pastor did hypothesize that both item and respondent characteristics would significantly affect endorsement of the midpoint such that the probability of item midpoint response selection would increase: (1) as ordinal item position increased, (2) if the item is reverse scored, (3) as respondent verbal ability (SAT verbal score) decreased, (4) as respondent self-reported effort decreased, and (5) as respondent self-reported perceived importance decreased. It should be noted that Marsh and Pastor intentionally did not specify any *a priori* hypotheses

regarding the effect of midpoint response option label on respondent endorsement of the midpoint response due to the exploratory nature of Study 1.

The results of Marsh and Pastor's (2011) Study 1 indicated that item ordinal position, item midpoint response option label, and respondent levels of self-reported effort were the only significant predictors of midpoint response option endorsement. Interestingly, the effect for item position was somewhat counterintuitive, as probability of midpoint response endorsement decreased significantly as item ordinal position increased. In addition, the probability of midpoint response endorsement increased for items with no midpoint response label versus those items with a midpoint response label of either *neutral* or *neither agreement nor disagreement*. Finally, the effect for respondent effort was congruent with Marsh and Pastor's hypothesis that the probability of midpoint response endorsement would increase as respondent levels of effort decreased. The relative predictive utility or importance of the significant item and respondent characteristics in predicting midpoint response endorsement was assessed by comparing the standardized odds ratios associated with each predictor. The strongest of these significant predictors was student level of self-reported effort, followed by whether or not the midpoint response option was labeled or unlabeled, and lastly, item ordinal position.

Preliminary results from the Marsh and Pastor (2011) study prompted the researchers to attempt to replicate and build upon the findings from Study 1. Marsh and Pastor (2012) sought to improve upon the previous study by including additional item level predictors and expanding the item pool to reflect more diverse item characteristics. Analyzing data from the same assessment session but a different subsample of students

completing alternate noncognitive assessments, the item pool was expanded to include more unrelated items from conceptually distinct noncognitive instruments with various response scales and midpoint response option label variations. Item level predictors of midpoint response option endorsement included within Study 2 were: item instrument-specific ordinal position (e.g., ordinal position of an item on a given instrument), item assessment session configuration-specific ordinal position (e.g., ordinal position of an item within the entire testing session), item reverse scoring (e.g., whether or not an item was reverse scored), item midpoint response option label (e.g., coded appropriately to reflect one of seven midpoint response option labels: *neutral*, *neither agreement nor disagreement*, *neither agree nor disagree*, *occasionally*, *occasionally characteristic of me*, *neutral/undecided*, or no label). It should be noted that the additional item assessment session configuration-specific predictor is not perfectly correlated with item instrument-specific ordinal position. This is because students within the current sample were randomly assigned to different assessment session testing configurations consisting of a different number and administration order of assessment instruments. Person level variables included as predictors of item midpoint response endorsement were identical to those included within the model in Study 1.

The results of Marsh and Pastor (2012) Study 2 indicated that item instrument position, item testing session position, item reverse scoring, all item midpoint response option labeling variations, and respondent self-reported effort were significant predictors of midpoint response option endorsement. Interestingly, the effects of item instrument position on midpoint response endorsement differed from those of item testing session position. More specifically, results suggested that as instrument item number increases,

the likelihood of midpoint response endorsement increases; whereas as testing session item number increases, the likelihood of midpoint response endorsement decreases, mirroring the findings of the item ordinal position variable in Study 1. In contrast to Study 1, in Study 2 the effect of item reverse scoring was significant such that respondents were more likely to endorse the midpoint response option when an item was reverse scored. The effects of all midpoint response option label variations were significant such that likelihood of midpoint response endorsement increased for each midpoint label variation (excluding the neutral midpoint label) compared to items with no midpoint response option label. For the neutral midpoint option label, the likelihood of midpoint response option endorsement decreased. Finally, the effect of respondent self-reported effort was replicated from Study 1, such that as levels of effort decreased, midpoint response option endorsement increased. Lastly, the relative predictive utility or importance of the significant item and respondent characteristics in predicting midpoint response endorsement was assessed again in Study 2 by comparing the standardized odds ratios corresponding to each predictor. The strongest of these significant predictors was the midpoint response option label *neither agree nor disagree* and the predictor with the least utility was *neutral/undecided*.

Given the paucity of research examining the effects of item and respondent characteristics on midpoint response endorsement in a noncognitive assessment context and somewhat inconsistent nature of the findings from the preliminary Marsh and Pastor (2011, 2012) studies, there remains a need to further investigate factors that potentially influence midpoint response abuse and respondent conceptualization of the midpoint. Though the results of Marsh and Pastor studies are promising in that they suggest

endorsement of the midpoint response option may not depend solely upon respondent levels of the construct of interest, there are a number of potential confounds present within their design that limit the inferences that can be made based upon their findings. The majority of the limitations present in the Marsh and Pastor studies stem from the fact that the data analyzed were not collected for the purpose of studying midpoint response option abuse or midpoint endorsement behavior. Instead, due to the exploratory nature of this initial research, the data were originally collected for higher education accountability purposes and then analyzed post hoc by the researchers. Although samples of noncognitive assessment items included within both studies were selected in such a way as to decrease the likelihood of consistent responding across items, this approach does not ensure that these items are completely heterogeneous in content and the constructs they measure.

Furthermore, the post hoc nature of these analyses results in the confounding of several item characteristics included as item-level predictors within the model. More specifically, because Marsh and Pastor (2011, 2012) did not have control over how the items were administered to respondents, item characteristics such as item ordinal position (both instrument- and testing session-specific) and item midpoint response option label were confounded with specific noncognitive assessment scales used. This poses an issue when interpreting these effects as researchers are unable to disentangle the effects of these item characteristics on midpoint response endorsement from item content. To better illustrate this issue, consider the findings of Marsh and Pastor (2011) Study 1 where the effects of item ordinal position and item midpoint response option label were significant. As item ordinal position increased, probability of midpoint response option endorsement

decreased. Because items were administered in order by noncognitive assessment scale, we cannot be sure that this increased probability of midpoint response option endorsement is not due to the nature of the constructs administered later within the assessment measure. Findings from Study 1 also suggested that respondents were less likely to endorse the midpoint response option when it had a label (either *neutral* or *neither agreement nor disagreement*) versus none. However, these labeling conventions were consistent within each noncognitive assessment scale. This is cause for concern as one cannot distinguish the effects of these variables on midpoint response option endorsement (e.g., is likelihood of midpoint response option endorsement affected by presence of midpoint response option label or due to a true difference in middling levels of the construct across scales). An ideal study for examining midpoint response endorsement behavior would utilize items unrelated in content, randomize the administration order of items and midpoint response option label across items allowing the effect of these characteristics to be studied without confound.

An additional potential confound associated with the item-level predictors used in the Marsh and Pastor (2011, 2012) studies discussed above relates to the effect of midpoint response option label. In both studies, the number of items with a midpoint response option label or associated with a specific midpoint response option labeling convention is rather small. For instance, within Marsh and Pastor Study 1 in which the effect of midpoint response option label was significant, only 6 of the 30 items utilized within the analyses actually had a midpoint response option label. Similarly, in Study 2, items within each of the six midpoint response option label categories were also rather small (with all but one category consisting of 5 or fewer items). These severely limited

item-grouping sample sizes result in heightened standard error estimates for these effects, calling into question the reliability and stability of these effects on midpoint response option endorsement. Ultimately, this issue paired with the ones mentioned above may be the source of some of the major discrepancies between the findings from the two Marsh and Pastor studies.

Research Questions

The purpose of the current study was to expand upon this line of research and improve upon the experimental design employed in the Marsh and Pastor studies to further examine item and respondent characteristics that potentially influence midpoint response abuse in a noncognitive assessment context. These effects were further examined within this dissertation by way of two studies – one quantitative (Study 1) and one qualitative (Study 2). The quantitative study serves as an attempt to expand upon the findings from preliminary quantitative approaches employed by Marsh and Pastor (2011, 2012) described above. In order to improve upon the approaches taken by Marsh and Pastor, item set location and midpoint response option label were experimentally manipulated across randomly equivalent groups of respondents administered the same two noncognitive assessment scales. These noncognitive assessments, developed to measure two distinct constructs, were consistently administered across randomly formed respondent groups; only item set location and midpoint response option label differed across these groups. Because the groups of respondents were randomly assigned, they should not differ systematically in their underlying levels of the construct measured by each set of noncognitive items, limiting the potential confound for item content to drive midpoint response option endorsement as noted in the Marsh and Pastor studies. Thus,

any significant differences in probability of midpoint response option endorsement across these groups should be attributable to these experimentally manipulated item characteristics or differences in respondent characteristics across groups. This approach will allow stronger claims regarding the impact of these item characteristics on midpoint response option endorsement, possibly bolstering evidence for respondent midpoint abuse in practice and the causal link between item predictors and this kind of response behavior.

Study 1. Based upon the preliminary findings of Marsh and Pastor (2011, 2012) and the experimental manipulations employed within the current design, research questions for Study 1 include:

- (1) Does experimental manipulation of item set location (earlier versus later) within an assessment instrument significantly affect midpoint response endorsement?
- (2) Does experimental manipulation of item midpoint response option label significantly affect probability of midpoint response option endorsement?
- (3) Do respondent self-reported levels of effort expended on assessments completed for higher education accountability purposes significantly affect probability of midpoint response endorsement?
- (4) Do respondent levels of verbal aptitude significantly affect the probability of midpoint response option endorsement?
- (5) Do these effects (item set location, midpoint response option label, and respondent levels of effort and/or verbal aptitude) significantly interact to affect probability of midpoint response option endorsement?

A hierarchical generalized linear modeling approach was utilized in order to examine the effects of these experimentally manipulated item and respondent characteristics on midpoint response endorsement. This approach simultaneously estimates the effects of item and respondent dispositional characteristics hypothesized to influence midpoint response endorsement, allowing the researcher to assess the unique predictive utility of each variable included within the model as well as their interaction effects. Moreover, a hierarchical generalized linear modeling approach allows the researcher to control for dependency of participant item responses by modeling participant noncognitive assessment item responses as nested within each respondent. Although there are two levels of analysis within the series of models utilized in Study 1, all item and respondent characteristic predictors and their interaction effects were modeled within Level 2 including: item set location (e.g., administration order of the two noncognitive assessments – earlier within the assessment instrument administered to respondents versus later), systematic midpoint response option label variation administered to the respondent (e.g., no midpoint response option label, midpoint response option labeled *neutral*, midpoint response option labeled *undecided*, or midpoint response option labeled *neither agree nor disagree*), respondent self-reported level of effort put forth on assessment instruments completed, and respondent verbal aptitude (i.e., SAT verbal score).

According to Krosnick's (1999) theory of satisficing versus optimizing, there may be reason to believe that variables such as item set location, respondent levels of effort expended on assessments completed during the assessment testing session, and respondent verbal aptitude have the potential to affect probability of midpoint response

endorsement in predictable ways. That is, respondents may be more or less prone to engage in satisficing behavior that manifests in the form of construct-irrelevant midpoint response option endorsement under certain conditions of these variables. Recall that Krosnick describes conditions that foster satisficing, specifically noting that respondent levels of cognitive effort expended on an item largely influence propensity to satisfice versus optimize in response option selection. When respondents exert less cognitive effort than is required to optimize, they are said to revert to satisficing.

If cognitive effort does play a role in response option selection and satisficing behavior that potentially manifests in midpoint response option abuse, we would expect that predictors such as item set location, respondent levels of effort expended on assessments, and respondent verbal aptitude may predict midpoint response option endorsement. For this reason, the hypotheses for each Study 1 research question related to these predictors are as follows:

- (1) As item set location is experimentally manipulated so that noncognitive item sets are administered later rather than earlier within an assessment instrument, probability of midpoint response option endorsement will increase. When items are administered later versus earlier, respondents will be more likely to experience response fatigue potentially diminishing levels of cognitive effort exerted by the respondent.
- (2) As respondent levels of self-reported effort expended on assessments completed during the assessment testing session decrease, probability of midpoint response option endorsement will increase. When respondents report

lower levels of effort expended on these assessments they will be less likely to exert the cognitive effort required to optimize their response selection.

- (3) As respondent verbal aptitude decreases, probability of midpoint response option endorsement will increase. Respondents exhibiting lower levels of verbal aptitude may be less likely to comprehend item statements and thus be unable to exert the cognitive effort required to optimize their response selection.

With regard to midpoint response option label, there was no definitive hypothesis. Although it is reasonable to believe that this variable will have some kind of meaningful effect on probability of midpoint response option endorsement based upon the preliminary results of the Marsh and Pastor (2011, 2012) studies and logical differences in meaning associated with each labeling convention as suggested by DeMars and Erwin (2005), there is no definitive evidence to inform how midpoint response option endorsement will be impacted by the experimental manipulation of this variable. Furthermore, there are no definitive hypotheses associated with the interaction effects amongst each variable included within the series of models utilized in Study 1. The investigation of these interaction effects is novel and purely exploratory in nature.

Study 2. The qualitative approach for Study 2 expands upon the findings from preliminary quantitative approaches employed by Marsh and Pastor, by providing supplemental data to support or refute differences in respondent midpoint response option conceptualization based upon manipulation of the midpoint response option label.

Research questions for Study 2 include:

(1) Do respondents appear to abuse the midpoint in practice? Do they provide construct-irrelevant justifications for endorsing the midpoint in response to the items administered?

(2) Do respondents conceptualize the midpoint response option differently depending on systematic variations in the midpoint response option label?

A think-aloud protocol was utilized in order to provide a more thorough understanding of differences in the justification(s) respondents provide for midpoint response option endorsement and their conceptualization of this response option. This approach, which instructs participants to think aloud while responding to a selection of the noncognitive assessment items administered in Study 1 across which the midpoint response option label was systematically varied, will allow the researcher to collect and analyze qualitative data that either supports or refutes the claim that the midpoint response option is abused in practice and that its label potentially alters respondent conceptualization and endorsement of the midpoint response. Both research questions examined within Study 2 are exploratory in nature, thus there are no specified a priori hypotheses associated with either.

III. Methods

Chapter Overview

Data for both studies were collected during the same university-wide mandatory assessment testing session, or Assessment Day, held in August 2012. Participants in both studies consisted of incoming first-year students scheduled to complete a series of assessments for higher education accountability purposes on the Friday before the beginning of the fall 2012 semester at a mid-sized southeastern university. During these Assessment Day testing sessions, students are administered approximately 5 tests to complete during a 3-hour period. Trained proctors deliver standardized instructions and oversee the administration of the assessments. These tests are low-stakes for students meaning that their performance on these tests does not factor into grades or graduation status, but are high-stakes for the university as data are used to evaluate student learning outcomes. Accordingly student motivation to perform well on these assessments is consistently monitored via administration of a self-reported test-taking motivation measure (which includes both test importance and test-taking effort subscales) given at the end of each 3-hour testing session.

The 20 noncognitive assessment items used in Study 1, where item and respondent level variables were used to predict midpoint response option endorsement, were administered as part of one assessment students completed during this Assessment Day session known as the SDA-7. Study 1 participants were randomly assigned to take one of 8 forms of the SDA-7 with specific forms corresponding to different manipulations of midpoint response option label and item set location for the 20 items of interest on the test. A series of hierarchical generalized linear models (HGLMs) were

utilized to examine whether respondent characteristics and systematic manipulations in item set location and midpoint response option label were related to probability of midpoint response option endorsement.

For Study 2, the researcher conducted a series of think-alouds with eight students using a 12-item subset of the 20 items used in Study 1. Students participating in the think-alouds were administered an identical set of items associated with different midpoint response option labels and variations in item-group administration order. Think-aloud participants were asked to think aloud as they responded to the items in order to obtain a better understanding of how variations in midpoint response option label and other item characteristics might impact midpoint response endorsement. In addition, students were asked structured interview questions to provide supplementary information that may not have been provided unprompted by the participants during the think-aloud exercise. Study-specific details of methodologies employed in each are more fully elaborated upon below.

Study 1

Participants and Procedure

Participants were a subsample of incoming first-year students participating in the fall 2012 Assessment Day. During Assessment Day, students are randomly assigned by the last two digits of their university assigned student ID number to one of several testing rooms and any number of testing configurations administered in each room (see Table 1 for SDA-7 Fall 2012 Assessment Day testing configurations). Students included within the current study were randomly assigned to any one of these rooms administering the SDA-7. A total of 1,826 students were included within the analyses for the current study.

The majority of these students were female (62.32%) and White (81.16%), with a mean age of 18.43 years ($SD = 0.36$).

Measures

Sociocultural Domain Assessment Version 7 (SDA-7). During the fall 2012 Assessment Day administration, a random sample of students completed the 7th version of the Sociocultural Domain Assessment which includes both cognitive and noncognitive assessment components. The SDA-7 is a 77 item instrument comprised of 29 cognitive multiple choice items assessing student knowledge in the sociocultural domain of general education, followed by a series of noncognitive measures. There were two noncognitive measures utilized for the purposes of the current study: a revised 6-item version of the Lounsbury, Levy, Leong, and Gibson (2007) Sense of Identity Scale (SoI; Samonte, 2011, see Appendix A) and a revised 14-item version of the Miville et al. (1999) Miville-Guzman University-Diversity Scale (M-GUDS; Pastor & Cotten, 2009, see Appendix B).

The revised 6-item version of the SoI (Samonte, 2011) was reduced from the original 8-item Lounsbury et al. (2007) version of the scale that was developed to measure respondent self-reported sense of identity or knowledge of self and sense of purpose (Lounsbury, Huffstetler, Leong & Gibson, 2005). This reduced 6-item SoI has been shown to conform to a one-factor structure yielding a reliable total score ($\omega = .81$) when administered to a college student sample (Samonte, 2011). The revised 14-item version of the M-GUDS (Pastor & Cotten, 2009) was reduced from the original 45-item Miville et al. (1999) version of the scale that was developed to measure self-reported respondent universal-diverse orientation or “an attitude of awareness and acceptance of both the similarities and differences among people” (p. 291). The reduced 14-item M-

GUDS is the result of an exploratory factor analysis conducted by Pastor and Cotten (2009) in order to better understand the structural validity of scores from the original M-GUDS. Results of this exploratory factor analysis indicate a four-factor structure with 4 items associated with discomfort with racial diversity loading onto one factor (DRD), 3 items associated with self-understanding through knowledge about diverse others' experiences loading onto a second factor (SELFUND), 4 items associated with the extent to which one better understands people by learning how they are both similar and different from oneself loading onto a third factor (SIMDIF), and 3 items associated with ease and comfort with others who differ in their sexual orientation loading onto a fourth and final factor (ECSO).

Students completing the SoI and M-GUDS measures as part of the SDA-7 were prompted to indicate their level of agreement with each item statement using a Likert-type response scale ranging from 1-7 (see Table 2). Midpoint response option labels were systematically varied across four forms of the SoI and M-GUDS in order to examine the effects of midpoint response option label on probability of midpoint endorsement: (1) Midpoint unlabeled (e.g., anchor points labeled only, 1 "Strongly Disagree", 7 "Strongly Agree"), (2) *Neutral* midpoint (e.g., 1 "Strongly Disagree", 2 "Disagree", 3 "Slightly Disagree", 4 "Neutral", 5 "Slightly Agree", 6 "Agree", 7 "Strongly Agree"), (3) *Undecided* midpoint (e.g., 1 "Strongly Disagree", 2 "Disagree", 3 "Slightly Disagree", 4 "Undecided", 5 "Slightly Agree", 6 "Agree", 7 "Strongly Agree"), (4) *Neither agree nor disagree* midpoint (e.g., 1 "Strongly Disagree", 2 "Disagree", 3 "Slightly Disagree", 4 "Neither Agree Nor Disagree", 5 "Slightly Agree", 6 "Agree", 7 "Strongly Agree").

SDA-7 forms were created in such a way that response scale versions were held constant across both SoI and M-GUDS measures across all respondents completing a given form (i.e., midpoint response option was consistently unlabeled across SoI and M-GUDS measures in forms A and E). In addition, SDA-7 SoI and M-GUDS administration order was varied in such a way that SoI and M-GUDS were administered either first and second (immediately following the 29-item cognitive portion of the SDA-7) or third and fourth (immediately following the administration of two additional noncognitive measures administered after the 29-item cognitive portion). However, the SoI was consistently administered prior to the M-GUDS across forms. Variation of SoI and M-GUDS response scale option versions and SDA-7 administration order resulted in a total of eight different SDA-7 forms (A-H, see Table 2). All eight forms were spiraled to ensure random administration across forms within each SDA-7 testing room. Students were instructed verbally by the testing session proctor and in writing on the test to indicate which form they were completing in response to the first SDA-7 item.

Student Opinion Scale (SOS). Each student completed the Student Opinion Scale (SOS, Sundre & Moore, 2002; Thelk, Sundre, Horst & Finney, 2009) upon the completion of the testing session (see Table 1). The SOS is a 10-item self-report measure indicating student levels of test-taking motivation comprised of two subscales, Effort and Importance. Each SOS subscale is comprised of 5 items, with Effort items evaluating levels of student effort expended on the assessments and Importance items evaluating student levels of perceived importance of the test. Students were prompted to indicate their level of agreement with each SOS item on a Likert scale (1 “Strongly Disagree”, 2 “Disagree”, 3 “Neutral”, 4 “Agree”, and 5 “Strongly Agree”) in regard to the series of

tests they completed during Assessment Day. Subscale scores range from a low of 5 to a high of 25 and higher subscale scores indicate higher levels of effort and perceived importance. Separate subscale totals were calculated for each student and only Effort scores were included as a respondent characteristic variable in the model.

Item Characteristics

Item set location. Item set location was included within the analysis as a Level 2 item characteristic predictor of midpoint response endorsement. As mentioned previously, SoI and M-GUDS item set location was systematically varied across SDA-7 forms A-H such that they appeared earlier (immediately following the 29-item cognitive portion of the SDA-7 forms A-D, with the SoI beginning at item 31 and M-GUDS ending on item 50) or later (immediately following the administration of two additional noncognitive measures administered after the 29-item cognitive portion of the SDA-7 forms E-H, with SoI beginning at item 58 and M-GUDS ending on item 77). A dummy-coded variable was included within the analysis to indicate item set location. This dummy-coded variable took on a value of “1” to indicate later administration of the SoI and M-GUDS item set (on forms E-H) or “0” to indicate earlier administration (on forms A-D).

Midpoint response option label. A series of dummy-coded variables were included within the analysis as Level 2 predictors of midpoint response option endorsement to indicate variation in item set midpoint response option label. Four different variations of midpoint response option label were present across forms: (1) Midpoint unlabeled (forms A and E), (2) *Neutral* midpoint (forms B and F), (3) *Undecided* midpoint (forms C and G), and (4) *Neither agree nor disagree* midpoint

(forms D and H). These three dummy-coded variables were included within the analysis as a series of item characteristic predictors that took on a value of “1” to indicate a specific midpoint response label (e.g., *neutral*, *undecided*, or *neither agree nor disagree*) or “0” to indicate that midpoint response option was left unlabeled.

Respondent Characteristics

Student self-reported effort. Student SOS Effort subscale scores ranging from 5 to 25 were included within the analysis as a Level 2 respondent characteristic predictor of midpoint response option endorsement.

SAT verbal score (SATV). Student SAT verbal scores were obtained from university records and included as a Level 2 respondent characteristic variable in the model. These scores were available from a university student database and merged with Assessment Day data for the 1,826 students included within the analyses. If a student took the SAT on multiple occasions, the highest SAT verbal score was utilized within the analyses.

Data Analysis

In order to simultaneously examine the relationships between endorsement of midpoint response and these item and respondent characteristics, a series of hierarchical generalized linear models were utilized conceptualizing items (Level 1) as nested within respondents (Level 2). Respondent endorsement of midpoint response (MR) served as the dependent variable, while item and respondent characteristics served as the independent variables within their respective models at Level 2. It should be noted that item level predictors (typically included within Level 1) were included within Level 2 of the models in the current study as each respondent received one of eight assessment forms associated

with unique item characteristic predictor experimental manipulations (e.g., item set location, midpoint response option label). Thus specified item characteristics such as item set location and midpoint response option label were included as predictors at Level 2 of the models. A series of four models were fit to the data. The first model, which did not include any predictors, was utilized to capture the average endorsement of MR in the respondent population and the extent to which MR selection varied across respondents. Within the second, third, and fourth models, the effects of various item and respondent characteristics on endorsement of MR were examined (in addition to their interaction effects). These series of models are described in more detail below.

Model 1 - Unconditional Model. Prior to including independent variables, an unconditional model, with no predictors, was fit to the data. Specifically, the probability of person j endorsing the midpoint response (MR) on item i was modeled at Level 1 as:

$$P(MR_{ij} = 1) = \frac{\exp(\beta_{0j})}{1 + \exp(\beta_{0j})} \quad (1)$$

The Level 1 probability model can also be written in reference to the log-odds or logit of endorsing the MR:

$$\log \frac{P(MR_{ij} = 1)}{P(MR_{ij} = 0)} = \beta_{0j} \quad (2)$$

In Equations 2 and 3, β_{0j} represents the log-odds that person j will endorse the MR. Specifically, β_{0j} represents the typical MR endorsement of person j across all 20 items. At Level 2, the β_{0j} s are modeled as a function of a single fixed effect, γ_{00} , and a single random effect, u_{0j} :

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (3)$$

where γ_{00} represents the average log-odds of endorsing the MR in the respondent population. The average log-odds of MR endorsement in the respondent population can be converted to a probability by substituting γ_{00} for β_{0j} in Equation 2ⁱ. The random effect or error, u_{0j} , represents the offset of person j from the average log-odds. These Level-2 errors are assumed to be normally distributed with a mean of zero and a variance equal to τ_{00} ($u_{0j} \sim N(0, \tau_{00})$). The Level 2 error variance indicates the extent to which endorsement of the MR varies in the respondent population.

Model 2 - Item set location, midpoint response option label, and their interaction effects. The Level 1 model was specified as:

$$\log \frac{P(MR_{ij} = 1)}{P(MR_{ij} = 0)} = \beta_{0j} \quad (4)$$

and the Level 2 model as:

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + \gamma_{01} \text{Item Set Location}_j + \gamma_{02} \text{Neutral}_j + \gamma_{03} \text{Undecided}_j + \gamma_{04} \text{Neither Agree nor Disagree}_j \\ & + \gamma_{05} \text{Neutral} * \text{Item Set Location}_j + \gamma_{06} \text{Undecided} * \text{Item Set Location}_j + \gamma_{07} \text{Neither Agree nor Disagree} * \text{Item Set Location}_j + u_{0j} \end{aligned} \quad (5)$$

Thus, the effects of item set location on endorsement of the MR can be examined controlling for the effects of item MR option label and vice versa. Moreover, the inclusion of item set location and MR option label interaction effects allows the researcher to further examine the interplay of these variables and their effects on MR option endorsement (e.g., does probability of MR option endorsement change significantly across variations in item set location and MR option label combinations?).

Model 3 - Item set location, midpoint response option label, respondent self-reported effort, and their interaction effects. The Level 1 model was specified as:

$$\log \frac{P(MR_{ij} = 1)}{P(MR_{ij} = 0)} = \beta_{0j} \quad (6)$$

and the Level 2 model as:

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + \gamma_{01} \text{Item Set Location}_j + \gamma_{02} \text{Neutral}_j + \gamma_{03} \text{Undecided}_j + \gamma_{04} \text{Neither Agree nor Disagree}_j + \gamma_{05} \text{Effort}_j \\ & + \gamma_{06} \text{Item Set Location} * \text{Neutral}_j + \gamma_{07} \text{Item Set Location} * \text{Undecided}_j + \gamma_{08} \text{Item Set Location} * \text{Neither Agree nor Disagree}_j \\ & + \gamma_{09} \text{Item Set Location} * \text{Effort}_j + \gamma_{10} \text{Neutral} * \text{Effort}_j + \gamma_{11} \text{Undecided} * \text{Effort}_j + \gamma_{12} \text{Neither Agree nor Disagree} * \text{Effort}_j \\ & + \gamma_{13} \text{Item Set Location} * \text{Neutral} * \text{Effort}_j + \gamma_{14} \text{Item Set Location} * \text{Undecided} * \text{Effort}_j + \gamma_{15} \text{Item Set Location} * \text{Neither Agree nor Disagree} * \text{Effort}_j + u_{0j} \end{aligned} \quad (7)$$

Thus, the effects of item set location, MR option label, and levels of respondent self-reported effort on endorsement of the MR can be examined controlling for the effects of each additional predictor in the model. Moreover, the inclusion of item set location, MR option label, and effort two- and three-way interaction effects allows the researcher to further examine the interplay of these variables and their effects on MR option endorsement.

Model 4 - Item set location, midpoint response option label, respondent SAT verbal score, and their interaction effects. The Level 1 model was specified as:

$$\log \frac{P(MR_{ij} = 1)}{P(MR_{ij} = 0)} = \beta_{0j} \quad (8)$$

and the Level 2 model as:

$$\begin{aligned} \beta_{0j} = & \gamma_{00} + \gamma_{01} \text{Item Set Location}_j + \gamma_{02} \text{Neutral}_j + \gamma_{03} \text{Undecided}_j + \gamma_{04} \text{Neither Agree nor Disagree}_j + \gamma_{05} \text{SAT Verbal}_j \\ & + \gamma_{06} \text{Item Set Location} * \text{Neutral}_j + \gamma_{07} \text{Item Set Location} * \text{Undecided}_j + \gamma_{08} \text{Item Set Location} * \text{Neither Agree nor Disagree}_j \\ & + \gamma_{09} \text{Item Set Location} * \text{SAT Verbal}_j + \gamma_{10} \text{Neutral} * \text{SAT Verbal}_j + \gamma_{11} \text{Undecided} * \text{SAT Verbal}_j + \gamma_{12} \text{Neither Agree nor Disagree} * \text{SAT Verbal}_j \\ & + \gamma_{13} \text{Item Set Location} * \text{Neutral} * \text{SAT Verbal}_j + \gamma_{14} \text{Item Set Location} * \text{Undecided} * \text{SAT Verbal}_j + \gamma_{15} \text{Item Set Location} * \text{Neither Agree nor Disagree} * \text{SAT Verbal}_j + u_{0j} \end{aligned} \quad (9)$$

Thus, the effects of item set location, MR option label, and SAT verbal scores on endorsement of the MR can be examined controlling for the effects of each additional

predictor in the model. Moreover, the inclusion of item set location, MR option label, and SAT verbal two- and three-way interaction effects allows the researcher to further examine the interplay of these variables and their effects on MR option endorsement.

In addition to the four models described in detail above, the researcher specified another exploratory model that included both respondent characteristic predictors (self-reported effort and SAT verbal score) and all possible interactions among these predictors.

Interpretation of Results. The significance of the coefficients in each model were evaluated using an $\alpha = .05$. The practical significance of the results was examined by computing and graphing the modeled probability of MR endorsement for various predictor combinations and noting the reduction of the conditional Level 2 error variance across models.

Software. Each of the models was analyzed using the NLMIXED procedure available in the software program, SAS (Version 9.2; SAS Institute Inc., 2008). The Gauss-Hermite quadrature (Pinheiro & Bates, 1995) was utilized as the estimation method.

Study 2

Participants and Procedure

Consistent with Study 1, participants were a selection of incoming first-year students participating in the fall 2012 Assessment Day. A total of eight students participated in the think-aloud process for the current study. Four of the participants were female and four were male, the majority were White (63%), with a mean age of 18.63 years ($SD = 0.74$).

Study 2 participants consisted of four female and four male students randomly selected from a subset of incoming first-year students assigned to complete assessments in the largest SDA-7 assessment testing configuration room (demographic information and ID numbers of those assigned to this testing configuration were available prior to assessment day administration). At the beginning of each testing session (AM and PM), the researcher projected names of potential study participants and directions advising students on the list to speak with their testing session proctor prior to the beginning of the assessment testing session. The proctor provided students with a form informing these students that they had been selected to participate in an alternate activity that would fulfill their assessment requirement if they chose to do so and directed them to the researcher present within the testing session room. Four prospective participants were escorted from the larger testing session room to another classroom to receive further information regarding their opportunity to participate prior to the beginning of assessment testing in the larger room during both AM and PM sessions.

Once prospective participants arrived in the classroom, the study was described in further detail (see “General introductions/instructions for golden ticket holders, Appendix C) and they received consent forms to read and sign prior to participation (see Appendix D). Students were informed that the research study should take approximately 2.5 hours to complete and they would be removed from the classroom at different times to participate in individual think-aloud protocols lasting 20-30 minutes with the researcher. When students were not participating in think-alouds, they were completing and providing feedback on a series of pilot instruments first administered during the fall 2012

assessment day (see “Fall 2012 Assessment Day Pilot Testing Student Feedback” instructions, Appendix E) under the supervision of a test proctor.

Study 2 data collection commenced upon completion of participant consent forms. Students were removed individually to participate in the think-aloud in another classroom. Prior to their participation, they received general instruction on think-aloud protocol with the researcher providing them with an example of the process (see “Researcher Introduction/Practice/Instructions”, Appendix F). They were then asked to practice this on their own with several practice items provided by the researcher (see “Practice Questionnaire”, Appendix G). Once respondents felt comfortable with the think-aloud process, the researcher began audio-recording and official data collection. Upon completion of the initial think-aloud, the researcher asked each participant to participate in a brief structured interview consisting of two follow-up questions (see “Structured Interview Questions”, Appendix H).

Measures

A subset of 12 items were selected from the revised SoI (Samonte, 2011) and revised M-GUDS (Pastor & Cotten, 2009) and administered as part of the think-aloud protocol (SoI items 1-5, M-GUDS items 2, 4, 5, 7, 11, 12, and 14; see Forms A-H, Appendix I). M-GUDS items associated with discomfort with racial diversity (DRD) and ease and comfort with others who differ in sexual orientation (ECSO) were purposely excluded from the think-aloud protocol due to their sensitive nature. The 12 items were consistent across forms, but their placement was altered systematically across forms in order to avoid confounds due to survey item administration order resulting in eight different think-aloud protocol forms (A-H) – one unique to each participant.

Students participating in the think-aloud were prompted to indicate their level of agreement with each SoI or M-GUDS item statement using a Likert-type response scale ranging from 1-7. Three to four item subsets of SoI and M-GUDS items were associated with different response option labels: (1) Midpoint unlabeled (e.g., anchor points labeled only, 1 “Strongly Disagree”, 7 “Strongly Agree”), (2) *Neutral* midpoint (e.g., 1 “Strongly Disagree”, 2 “Disagree”, 3 “Slightly Disagree”, 4 “Neutral”, 5 “Slightly Agree”, 6 “Agree”, 7 “Strongly Agree”), (3) *Undecided* midpoint (e.g., 1 “Strongly Disagree”, 2 “Disagree”, 3 “Slightly Disagree”, 4 “Undecided”, 5 “Slightly Agree”, 6 “Agree”, 7 “Strongly Agree”), (4) *Neither agree nor disagree* midpoint (e.g., 1 “Strongly Disagree”, 2 “Disagree”, 3 “Slightly Disagree”, 4 “Neither Agree Nor Disagree”, 5 “Slightly Agree”, 6 “Agree”, 7 “Strongly Agree”). However, midpoint response option labels associated with each SoI and M-GUDS item subset remained constant across the forms (e.g., unlabeled midpoint was always the response option scale provided for SoI item 1 and M-GUDS items 12 and 14), thus the only variation across forms was administration order.

Data Analysis

Think-aloud protocol data for each participant was audio-recorded and transcribed by the researcher. Upon transcription, qualitative data was analyzed according to the thematic network analysis strategy described by Attride-Stirling (2001). This qualitative data analysis strategy unfolds in three stages and consists of six steps: Stage A: A reduction and breakdown of text; Steps: (1) Code material (2) Identify themes (3) Construct thematic networks; Stage B: Exploration of the text; Steps (4) Describe and explore thematic networks (5) Summarize thematic networks; Stage C: Integration of exploration; (6) Interpret patterns (Attride-Stirling, 2001, p. 391). During the initial stage

(“reduction and breakdown of text”), the researcher develops and applies a coding scheme to the transcribed data from each participant think-aloud, identifies appropriate themes to organize coded data, and creates a visual representation of this data organization via development of what Attride-Stirling calls a “thematic network”. The thematic network is a visualization of the organization of qualitative data codes across varying categories of themes, each increasingly broader to encompass more data and more succinctly describe overarching themes determined to emerge from the data. Once this thematic network is established and “refined”, the researcher may move onto the next stage of the qualitative data analysis strategy (“exploration of the text”), during which this newly developed thematic network facilitates interpretation of the text. At this point, the researcher utilizes the themes and associated segments of text to facilitate description and overall interpretation of the data in order to move onto the third stage of analysis (“integration of exploration”). This final stage culminates with the researcher’s attempt to marry the emergent themes and tie them back to initial research questions and appropriate theory.

IV. Results

Chapter Overview

A mixed-methods approach was employed in order to address a series of research questions posed to examine the potential for midpoint response option abuse in noncognitive assessment practice. The current section reports the Study 1 (quantitative) and Study 2 (qualitative) results separately, linking the findings of each study to their respective a priori research questions.

Study 1

Data Cleaning

A total of 2,403 students were administered the SDA-7 during the fall 2012 assessment day. By design, 285 of these students did not receive any one of the eight experimentally manipulated SDA-7 forms (A-H) and were removed prior to data analysis. Five more students were removed from the data set because they failed to indicate which SDA-7 form they had completed by responding to the first SDA-7 item, resulting in a total N of 2,113. Cases with missing data for variables included within all models ($n = 287$) were listwise deleted resulting in a final N of 1,826.

Descriptive Statistics

In order to summarize overall MR endorsement, the average endorsement of the MR by item across respondents and the average endorsement of the MR by respondent across items were calculated. Based on the average endorsement of the MR by item across respondents, for a typical item, approximately 13% of respondents endorsed the midpoint. However, there was sizeable variability among items in the proportion of respondents endorsing the midpoint, with proportions ranging from a low of 4% for item

3 (SoI item 3, “I have a clear set of personal values or moral standards”) and a high of 27% for items 10 (M-GUDS item 4 (SELFUND), “Knowing someone from a different ethnic group broadens my understanding of myself”) and 13 (M-GUDS item 7 (SIMDIF), “Knowing how a person differs from me greatly enhances our friendship”). The distribution of the average endorsement of the MR by respondent across items, or MRR, values are shown in Figure 1. On average, 13% of responses to the 20 SoI and M-GUDS items were at the midpoint. As can be seen in Figure 1, there is substantial variability in the average endorsement of the midpoint across respondents and the distribution is positively skewed.

Intercorrelations and descriptive statistics for variables included within Models 1-4 are shown in Table 3. Correlations among predictors exceeding $|.70|$ were interpreted to suggest potential problems with multicollinearity. Because no correlation exceeded $|.70|$ no multicollinearity problems were indicated. Descriptive statistics for additional Study 1 variables (SoI total score, M-GUDS subscale scores, and percent correct scores on the multiple-choice portion of the SDA-7) and variables included within Models 1-4 (SOS effort, SAT verbal score, and MRR) are shown within Table 4 for each SDA-7 form (A-H) experimental manipulation. Two-way ANOVA results assessing differences in all variables (except MRR) due to item set location, midpoint label or the interaction of item set location and midpoint label are also provided. ANOVA results are not provided for average MR endorsement because it is a Poisson distributed variable, as opposed to a normally distributed variable. Moreover, ANOVA results were also not reported for average MR endorsement because the effects of the experimental manipulations of item set location and midpoint label on MR endorsement are provided by Model 2.

The means on many of the variables are very similar for each group of respondents regardless of the SDA-7 form they completed. Two of the variables in Table 4, SAT verbal scores and the percent correct scores on the multiple-choice portion of the SDA-7, were collected prior to respondents encountering the experimental manipulation in each form. Therefore, the finding that average SAT verbal scores and average percent scores on the multiple-choice portion of the SDA-7 do not differ across the eight conditions provides evidence to support the claim that respondents completing each of the forms are randomly equivalent. These results thus strengthen our ability to attribute differences among groups in MR endorsement to the experimental manipulations of item set location and MR option label.

The remaining variables had the potential to be influenced by the experimental manipulations. These variables include the scores on the SoI and M-GUDS subscales, and the respondents' self-reported effort scores. Although some results were statistically significant, no practically significant average differences across the eight conditions were found for these variables. These findings indicate that the type of midpoint label used and the location of the item within the scale are not strongly related to total scores on each scale or respondents' self-reported effort during the assessment testing session.

Model 1 - Unconditional Model. The average log-odds of MR endorsement in the respondent population was estimated using the unconditional model to be $\gamma_{00} = -2.16$. This corresponds to an average probability of MR endorsement equal to 0.10, close to the observed probability of 0.13. The variance of MR endorsement across respondents was statistically significant ($t(1,825) = 16.48, p < .001$) and estimated as $\tau_{00} = 0.86$.

Model 2 - Item set location, midpoint response option label, and their interaction effects. The results for Model 2 are displayed within Table 5. In addition, the practical significance of these results was examined by computing and graphing the predicted probability of MR endorsement for various Model 2 predictor combinations (see Figure 2). Using the model significance values and an $\alpha = .05$, the only statistically significant effect in Model 2 was item set location. When interpreting the parameter estimates, it is important to recall that three dummy-coded variables were included within the model to reflect the various MR option label conditions (with each dummy-coded variable taking on a value of “1” to reflect either *neutral*, *undecided*, or *neither agree nor disagree*). The midpoint label condition in which the MR was not labeled, or the *unlabeled* condition, served as the reference group. The positive coefficient associated with item set location indicates that as item set location increases (i.e., when SoI, M-GUDS items were administered later in the SDA-7), probability of MR endorsement also increases for the midpoint label reference group, or the *unlabeled* condition. The lack of significance of the interaction terms indicates that all remaining MR option label conditions (*neutral*, *undecided*, or *neither agree nor disagree*) do not differ from the *unlabeled* condition in the difference between MR endorsement for items administered later versus earlier in the SDA-7.

The predicted probabilities of MR endorsement are very similar across all eight conditions. The only notable difference between predicted probabilities occurs in the *unlabeled* condition, where the probability of MR endorsement tended to increase when the 20 noncognitive assessment items were administered to respondents later (as items 58-77) in the SDA-7 versus earlier (as items 31-50). Although the item set location effect

appears to differ across midpoint conditions in Figure 2, with a larger difference for the *unlabeled* condition and essentially no difference in the remaining conditions, the statistical significance of this parameter estimate indicates that the significant item location effect in the *unlabeled* condition applies to all conditions.

The conditional Level 2 error variance in Model 2 was statistically significant ($t(1,825) = 16.43, p < .001$) and estimated as $\tau_{00} = 0.85$. This estimate was reduced compared to what was estimated in Model 1 (0.86), suggesting that the additional predictors and interaction terms included within Model 2 helped to explain some additional variance in MR endorsement amongst respondents; however, this reduction was very small.

Model 3 - Item set location, midpoint response option label, respondent self-reported effort, and their interaction effects. The results for Model 3 are displayed within Table 6. In addition, the practical significance of these results was examined by computing and graphing the predicted probability of MR endorsement for various Model 3 predictor combinations (see Figure 3). Using the model significance values and an $\alpha = .05$, the statistically significant effects in Model 3 included item set location, self-reported effort, and the *neither agree nor disagree* midpoint label and item set location interaction term. The negative coefficient associated with effort indicates that as levels of self-reported effort increase, the probability of MR endorsement decreases. Because effort does not interact with any other variable in the model, the negative relationship between effort and MR endorsement holds across all eight conditions.

The positive coefficient associated with item set location indicates that, controlling for effort, as item set location increases, probability of MR endorsement

increases for the midpoint label reference group, or the *unlabeled* condition. The lack of significance of the interaction terms for *neutral* and *undecided* with item set location indicates that the item set location effect in the *unlabeled* condition also applies to these conditions. Model 3 differs from Model 2 in that a significant interaction was found between the *neither agree nor disagree* MR option label and item set location. This effect indicates that after controlling for respondent effort, the effect of item set location for the *neither agree nor disagree* condition differs from the *unlabeled* condition. The predicted probabilities according to this model illustrating the difference in the item set location effect for these two conditions controlling for effort are almost identical to those found for the *unlabeled* and *neither agree nor disagree* conditions in Figure 2.

Although effort was entered into the model as a continuous variable, in Figure 3 predicted values are shown only for two levels of effort, corresponding to one standard deviation above (a value of about 22) and below (a value of about 15) the mean. The most notable effect seen in Figure 3 is associated with respondent effort. We see that probability of MR endorsement is lower for respondents reporting higher levels of effort, for all item set locations and midpoint labels. The significant item set location effect indicates that in all but the *neither agree nor disagree* MR label condition, respondents receiving the item set later exhibited a higher probability of MR endorsement than those receiving the item set earlier, regardless of differences in self-reported effort (low versus high). The item location effect significantly differs in the *neither agree nor disagree* MR label condition; where the item set location effect is reversed.

The conditional Level 2 error variance in Model 3 was statistically significant ($t(1,825) = 16.09, p < .001$) and estimated as $\tau_{00} = 0.79$. This estimate was reduced

compared to what was estimated in Models 1 and 2, suggesting that the additional predictors and interaction terms included within Model 3 helped to explain some additional variance in MR endorsement amongst respondents.

Model 4 - Item set location, midpoint response option label, respondent SAT verbal score, and their interaction effects. The results for Model 4 are displayed within Table 7. In addition, the practical significance of these results was examined by computing and graphing the predicted probability of MR endorsement for various Model 4 predictor combinations (see Figure 4). Using the model significance values and an $\alpha = .05$, the statistically significant effects in Model 4 included item set location and the interactions between the *neutral* and *undecided* midpoint label and SAT verbal scores. The positive coefficient associated with item set location indicates that as item set location increases, probability of MR endorsement also increases. Because item set location did not interact with any other variable in the model, the positive item set location effect applies across all conditions and SAT verbal score levels.

In order to understand the significant interactions between SAT verbal scores and two of the midpoint label conditions, the main effect for SAT verbal should be described. The near significant negative coefficient associated with SAT verbal score indicates that as levels of respondent verbal aptitude increase, probability of MR endorsement decreases for respondents in the *unlabeled* MR option label condition. The significant interaction effects between the *neutral* and *undecided* midpoint response option labels and SAT verbal score indicates that probability of MR endorsement significantly differs for these MR option labels compared to the *unlabeled* condition across levels of respondent verbal aptitude. This difference lies in the increased probability of MR

endorsement when respondents exhibit higher levels of verbal aptitude for these MR option label conditions.

Although SAT verbal score was entered into the model as a continuous variable, in Figure 4 predicted values are only shown for two levels of SAT verbal score, corresponding to one standard deviation above (a value of about 636) and below (a value of about 504) the mean. The significant item set location main effect indicates that the predicted probabilities are differentiated by item set location within each SAT level and MR label condition, such that respondents receiving the item set later more frequently endorsed the midpoint. The significant interaction between SAT verbal and the *neutral* and *undecided* MR option labels indicates that, controlling for item set location, the difference between low and high SAT verbal respondents in endorsing the midpoint differs between these conditions and the *unlabeled* condition. Whereas low SAT verbal respondents are more likely than high SAT verbal respondents to endorse the midpoint in the *unlabeled* condition, the opposite is true in the *neutral* and *undecided* conditions.

The conditional Level 2 error variance in Model 4 was statistically significant ($t(1,825) = 16.32, p < .001$) and estimated as $\tau_{00} = 0.83$. This estimate was reduced compared to what was estimated in Models 1 and 2, suggesting that the additional predictors and interaction terms included within Model 4 helped to explain some additional variance in MR endorsement amongst respondents. However, the reduction in τ_{00} in Model 4 is not as sizeable as the reduction of τ_{00} in Model 3, perhaps suggesting that SAT verbal score (and its interactions with other predictors) may not account for as much variation in MR endorsement as respondent self-reported effort (and its interactions with other predictors).

Models 5 and 6 - Item set location, midpoint response option label, respondent self-reported effort, SAT verbal score, and their interaction effects. The results for exploratory Model 5 are displayed within Table 8. Using the model significance values and an $\alpha = .05$, the statistically significant effects in Model 5 included item set location, effort, the *neither agree nor disagree* midpoint label and item set location interaction effect, the *neutral* and *undecided* midpoint label and SAT verbal score interaction effects, the *neutral* midpoint label, effort, and SAT verbal score interaction effect, and *neutral* midpoint label, item set location, effort, and SAT verbal score interaction effect.

To simplify interpretation, the significant effects from this model (and all lower order terms for interactions) were retained and included in Model 6. Most effects significant in Model 5 remained significant in Model 6, with the exception of the main effect for item set location and its interaction with the *neither agree nor disagree* midpoint label. The results of Model 6 are shown in Table 9 and the predicted probabilities according to the model for various combinations of the predictors are shown in Figures 5 and 6 for respondents with high and low SAT verbal scores, respectively.

The significant main effect for effort indicates that respondents with lower effort scores have a higher probability of endorsing the midpoint. A comparison of the solid lines (high effort) with the dashed lines (low effort) in Figures 5 and 6 clearly conveys the main effect for effort. In three of the four midpoint label conditions, higher reported effort was associated with lower midpoint endorsement, and this result did not depend on item set location or the SAT verbal score level of the respondent. For the *neutral* condition, however, the relationship between effort and midpoint endorsement depends

on item set location and the SAT verbal score of the respondent. This complicated interaction for the *neutral* condition appears to be driven by the difference in midpoint endorsement between high and low SAT verbal respondents in the early condition who report low effort. High SAT verbal respondents have a predicted probability of endorsing the midpoint equal to 0.19, whereas low SAT verbal respondents have a predicted probability equal to 0.09. The remaining noteworthy effect is the interaction between the *undecided* condition and SAT verbal. This interaction indicates that, controlling for all other variables (e.g., effort, item set location), the difference between low and high SAT verbal respondents in endorsing the midpoint differs between the *unlabeled* and *undecided* conditions. Whereas low SAT verbal respondents are more likely than high SAT verbal respondents to endorse the midpoint in the *unlabeled* condition, the opposite is true in the *undecided* condition.

The conditional Level 2 error variance in Model 6 was statistically significant ($t(1,825) = 15.99, p < .001$) and estimated as $\tau_{00} = 0.78$. This estimate was reduced compared to what was estimated in Models 1-4, suggesting that the additional predictors and interaction terms included within Model 6 helped to explain some additional variance in MR endorsement amongst respondents. However, this estimate was not drastically reduced compared to what was estimated in Model 3, perhaps suggesting that the inclusion of SAT verbal and its associated interaction terms in Model 6 may not account for a practically significant greater amount of variation in MR endorsement.

Study 1 Summary

Models 1-4 and an exploratory Model 5 were specified prior to data analysis to address the five research questions associated with Study 1. A sixth post-hoc model

including all significant effects from Model 5 (and retaining lower order terms for significant interaction terms) was specified and interpreted in addition to each of the other pre-specified models. In order to facilitate interpretation of Study 1 results, τ_{00} and fit indices were examined for Models 2 through 6 (see Table 10). Model 5 is the most complex model and as such, yielded the best fit to the data relative to all other models. A series of likelihood ratio tests were conducted comparing the fit of each of the constrained models (Models 1-4, Model 6) to Model 5 (which included all possible predictors and interaction terms specified in the constrained models). The results indicated that Models 2 through 4 fit significantly worse than Model 5. Model 6, however, did not significantly differ in fit compared to Model 5, indicating that this more parsimonious model should be favored. The information criteria (AIC, BIC), which penalize the deviance statistic in different ways for model complexity, were lowest for Model 6, also indicating retention of that model. Because Model 3 and Model 6 are similar in values of τ_{00} and deviance, Model 3 was also considered as it is more parsimonious than Model 6 and appears to explain about the same amount of between-respondent variability in MR endorsement. These models are not nested and therefore could only be compared using information criteria, which slightly favored Model 6 over Model 3. Based upon evaluation of information criteria and the results of the likelihood ratio tests shown in Table 10, Model 6 emerged as the favored model and was selected to be interpreted in more detail with regard to each of the research questions associated with Study 1.

Research Question 1: Does experimental manipulation of item set location (earlier versus later) within an assessment instrument significantly affect midpoint

response endorsement? Yes, given the results of Model 6, it appears that the experimental manipulation of item set location does significantly affect probability of MR endorsement. Although the main effect of item set location was no longer statistically significant in Model 6 ($p = 0.07$) as it was in Models 2 through 5, its positive value indicates that items administered later tend to result in increased MR endorsement. The *neutral* midpoint label condition, item set location, effort, and SAT verbal four-way interaction effect remained significant in Model 6. As previously explained, in the *neutral* midpoint label condition, the relationship between MR endorsement and effort depends on item set location and levels of SAT verbal. More specifically, this significant interaction effect appears to be driven by the variation in MR endorsement across levels of SAT verbal (low versus high) when respondents report lower effort and receive the items earlier in the SDA-7 versus later. In the *neutral* midpoint label condition, these predictors interact in such a way that respondents with higher SAT verbal scores actually exhibit the highest probability of MR endorsement.

Research Question 2: Does experimental manipulation of item midpoint response option label significantly affect probability of midpoint response option endorsement? Maybe; given the results of Model 6, while MR option label appears to interact with various item and respondent characteristics in unique ways to influence probability of MR option endorsement, there is no significant main effect of MR option label. Thus although the significance of the two-way *undecided* midpoint label and SAT verbal interaction term and the four-way *neutral*, item set location, effort, and SAT verbal interaction term suggest that MR endorsement significantly differs across these variations in midpoint label in comparison to the *unlabeled* condition, there is no evidence to

suggest that variations in MR option label consistently affects MR endorsement regardless of effort and SAT verbal across the eight experimental conditions.

Research Question 3: Do respondent self-reported levels of effort expended on assessments completed for higher education accountability purposes significantly affect probability of midpoint response endorsement? Yes, again, given the significant main effect of effort in Model 6, levels of respondent self-reported effort do appear to significantly affect MR endorsement. More specifically, as respondent self-report levels of effort decreased, probability of MR endorsement increased across the *unlabeled*, *undecided*, and *neither agree nor disagree* midpoint label conditions regardless of item set location or SAT verbal. Effort was also related to MR endorsement in the *neutral* condition, although its effect was dependent on item set location and SAT verbal score.

Research Question 4: Do respondent levels of verbal aptitude significantly affect the probability of midpoint response option endorsement? Yes, although there was no significant main effect of SAT verbal in Model 6, significant interaction terms included the two-way interaction between the *undecided* midpoint label condition and SAT verbal, and the four-way interaction amongst the *neutral* midpoint label, item set location, effort, and SAT verbal. These significant interaction terms suggest that probability of MR endorsement differs for these midpoint label conditions (*neutral* and *undecided*) compared to the *unlabeled* condition across levels of SAT verbal.

Research Question 5: Do these effects (item set location, midpoint response option label, and respondent levels of effort and/or verbal aptitude) significantly interact to affect probability of midpoint response option endorsement? Yes, given the series of significant interaction terms in Model 6 (each cited in response to the

previous research question), item set location, MR response option label, respondent self-reported levels of effort, and SAT verbal scores do appear to interact in such a way as to significantly affect probability of MR endorsement. More specifically, the significance of the highest order, four-way interaction of the *neutral* midpoint label, item set location, effort, and SAT verbal suggests that each of these respondent characteristics and experimentally manipulated item characteristics have the potential to interact in unique ways.

Study 2

Once think-aloud data were transcribed, student responses pertaining to MR option endorsement were coded and emergent themes were established by way of Attride-Stirling's (2001) thematic network analysis. Recall that participants were asked to participate in the think-aloud protocol primarily to address the first research question associated with Study 2 regarding the nature of respondent use of the midpoint response option: Do respondents appear to abuse the midpoint in practice? Do they provide construct-irrelevant justifications for endorsing the midpoint in response to the items administered? Following the think-aloud, the researcher asked the participants to respond to two structured interview questions (see "Structured Interview Questions", Appendix H) included within the study to address the second research question: Do respondents conceptualize the midpoint response differently depending on systematic variations in the midpoint response option label? The findings associated with both qualitative inquiries are discussed below.

Study 2 Research Question 1: Think-aloud Results

All student responses pertaining to midpoint endorsement were compiled into one master document for review. As each of the transcribed statements was reviewed by the

researcher, a corresponding label or statement believed to reflect core sentiments expressed by each respondent was provided. Upon this initial coding of think-aloud data, derived codes were reanalyzed and organized into broader categories reflecting a number of “basic themes” that emerged (Attride-Stirling, 2001). These basic themes were then further organized into broader categories referred to as “organizing themes” that ultimately relate to the “global theme” of interest – factors that influence midpoint response option endorsement (see Table 11) (Attride-Stirling, 2001). There were a total of five basic themes derived from the think-aloud data coding: *ambivalence*, *non-commitment*, *uncertainty*, *indifference*, and *item statement clarity*. From these basic themes, one broader organizing theme labeled *indecision* was established to encompass three of the most similar basic themes, *non-commitment*, *uncertainty*, and *indifference*. Basic themes reflecting respondent statements of ambivalence or issues with item statement clarity were considered broad enough that the researcher thought it would be inappropriate to create additional organizing themes.

Ambivalence. One emergent theme of respondent midpoint endorsement justification was ambivalence. As Breckler (1994) stated, “ambivalence (is) expressed when a person endorses both positive and negative attitudinal positions”, thus this basic theme encompassed any respondent statements of MR option endorsement that reflected feelings of agreement and disagreement in response to an item statement (p. 350). One respondent endorsing the midpoint and expressing ambivalence towards an item (“Knowing about the experiences of people of different races increases my self-understanding”) said, “I would choose four because I don’t think it really has a beneficial or negative impact; it just is what it is”. Similarly, another respondent expressed a similar

sentiment in response to an alternate item (“Knowing how a person differs from me greatly enhances our friendship”) stating, “Four - I neither agree nor disagree. Because some connections may be positive, some connections may be negative. I guess it really just depends on what those differences are...”. In general, both respondents endorsing the midpoint and expressing ambivalence towards these items expressed certainty in their response selection and specifically noted a kind of middling position reflective of a “leveling out” of agreement and disagreement with these item statements.

Indecision. The one organizing theme derived from the basic themes was *indecision*. This organizing theme encompasses three of five basic themes deemed to reflect similar respondent sentiments for endorsing the MR option for any given item, including *non-commitment*, *uncertainty*, and *indifference*. Although respondent justification for midpoint endorsement was similar across these three basic themes as they reflected general response selection indecision, they remained distinct enough to warrant independent coding and description.

Non-commitment. Several think-aloud participants appeared to be noncommittal in their justification provided for MR option endorsement. These respondents selected the MR option, but expressed either slight agreement or disagreement when justifying their response selection. One such respondent expressing slight agreement with an item statement (“Knowing someone from a different ethnic group broadens my understanding of myself”) explained, “Four – I’m neutral. I do like to meet people from different ethnic groups, like I had a best friend that’s from...I do like meeting from different ethnic groups. It does kind-of expand my horizons to know more things”. Thus, although this respondent selected the midpoint, this explanation indicated agreement with the

statement. Conversely, another respondent expressed slight disagreement with an item statement (“Knowing how a person differs from me greatly enhances our friendship”) while still endorsing the midpoint, “I neither agree nor disagree because whether I know how they’re different from me doesn’t really help me get along with them, it actually would have a negative impact”. Again, although this respondent selected the midpoint, this explanation indicated disagreement with the item statement.

Uncertainty. In addition to sentiments of non-commitment, some respondents expressed feelings of uncertainty about the applicability of an item statement (or part of an item statement) and how it pertains to them when endorsing the MR option. For example, one respondent noted the abstract nature of a particular item statement (“Knowing someone from a different ethnic group broadens my understanding of myself”) saying, “I would have to pick number four because I don’t know how it relates to myself or how it doesn’t – it seems good if you were to look at people in general and how they react to each other because of different ethnicities...to look at it as a part of yourself is kind of abstract”. Another participant noted the potential for inapplicability of an item statement (“Knowing about the experiences of people of different races increases my self-understanding”) when endorsing the midpoint stating, “I just, I don’t feel really strong that I know about my self-understanding, I don’t (know) that many experiences. But I do know some...”. Both responses appear to reflect a lack of connection with an item statement (or part of an item statement) that leads the respondent to feel that it is not entirely applicable to endorsement of the MR option.

Indifference. A number of respondents expressed indifference or a lack of opinion when endorsing the MR option for a given item. One respondent originally selected the

midpoint for an item (“Knowing about the experiences of people of different races increases my self-understanding”), but throughout the process of explaining the response selected a different option. When prompted to explain why he had originally selected the midpoint he said, “Cause I was kind of a little bit indifferent, but then once I said started talking about like knowing other people, like learning from them, it would help you better understand yourself through learning from other people”. An additional participant responding to the same item also noting indifference stated, “Kinda, but also kinda not. We’ll say four because I’m kind of indifferent about that one”. Both participants specifically noted a feeling of indifference towards the item statement when justifying endorsement of the midpoint, suggesting that respondents exhibiting indifference or lack of opinion towards a given item statement have the potential to revert to midpoint endorsement when a more appropriate option (like *no opinion*) is not provided.

Item statement clarity. Finally, there was one additional reoccurring justification for MR option endorsement provided by a few think-aloud participants that related to their perception of item statement clarity. More specifically, a number of respondents noted confusion (or lack of understanding as to what a given item was asking) when endorsing the MR option. For example, one respondent rationalized midpoint endorsement in response to one item (“Knowing someone from a different ethnic group broadens my understanding of myself”) by saying, “I didn’t necessarily agree or disagree, and at the same time I didn’t completely understand what it’s asking for”. Another also noted confusion or lack of item statement clarity in response to an alternate item (“I don’t know where I fit in the world”) mentioning, “Not entirely sure what this one is asking. I think maybe I know where I fit in the world, but I also don’t...so I’ll say four – neither

agree nor disagree, I suppose". When respondents expressed confusion or issues with item statement clarity, they appeared to simply default to midpoint endorsement.

Study 2 Research Question 2: Structured Interview Question Results

Although think-aloud data were analyzed primarily to address the first research question associated with Study 2, there were some findings that may supplement respondent data from the structured interview questions. Based upon think-aloud midpoint endorsement data, there seems to be little to no evidence to support systematic variation in MR option endorsement across label variations (i.e., *unlabeled*, *neutral*, *undecided*, *neither agree nor disagree*). More specifically, the frequency of think-aloud participants endorsing the MR option did not vary substantially across three of the four MR option label conditions (see Table 12). Four participants endorsed the MR for any one item within the grouping associated with *neither agree nor disagree* and five participants endorsed the MR for any one item within the grouping associated with the *unlabeled* and *neutral* conditions. Only one participant endorsed the MR for an item associated with the *undecided* label, however this cannot be attributed to MR option label variation alone as the item grouping and thus item statement content is unique to this label. In further support of this lack of systematic variation in MR option endorsement across label variations, respondents did not appear to offer consistent midpoint endorsement justifications for each midpoint label. In other words, respondents did not offer consistent rationalizations for endorsing the MR option for any one labeling convention. Instead, the nature of midpoint endorsement justification statements for think-aloud participants varied both within item grouping with the identical MR option labels and across item groupings with different label variations.

Following the think-aloud protocol, respondents were asked to respond to two structured interview questions. The first of these was intended to identify any participants who failed to notice the change in MR option labels across think-aloud item groupings. Two of eight think-aloud participants responded negatively to this first structured interview question with one respondent indicating that the response scales were, “pretty much the same” and another noting that the response scales changed but describing the change incorrectly noting that scale poles alternated, which was not the case. Data for the latter participant was retained for the second structured interview question after clarification regarding the change in MR option label, but not for the participant indicating no difference across the response scales. Of the six respondents that provided a positive response to the first structured interview question, four noted difference in MR option label and two more noted a change in response scale options.

The second structured interview question was a follow-up to the first, prompting participants to indicate whether they believed change in MR option label influenced how they responded to item statements. If a participant first indicated the possibility of MR option label influencing response selection, the participant was then prompted to explain specifically how this affected the response. Of the seven remaining participants, four indicated that systematic variation of the MR option label made no impact on their response selection. Two of these respondents indicated that they either had an opinion about the item statement or did not, thus midpoint label did not have a chance to play a role in their response selection. Another respondent indicated that midpoint label did not matter because, “You still know that the four is like, your in-between, and then as you go down from four it’s more of a disagree, and then, go up, it’s gradually higher – agree”

suggesting consistent interpretation of the midpoint (as a middling point on a continuum) regardless of variation in label. The final participant to indicate that midpoint label did not systematically influence response selection noted that she tended to stay away from the MR option, “just because the middle ones don’t do anything for data”. When prompted to expand upon this, she indicated that she would endorse the midpoint response option if she were “honestly at the midpoint” but would opt for another response option whenever appropriate. When asked to expand upon when she decides to endorse the MR option and when she opts for another more appropriate response this participant noted that she typically endorse the midpoint when, “(it) doesn’t matter” or she lacks an opinion in response to the item statement.

The three respondents indicating that variation in MR option label influenced their response selection were similar in a few notable ways. First, they were more likely to endorse the midpoint in the think-aloud portion of Study 2 in comparison to the other participants, each endorsing the midpoint on three occasions. Secondly, each of these participants mentioned a preference for response option labels (in general, across all response options). They all indicated that such labels provide more information and/or clarification for them when responding to an item and selecting their response. Thirdly, they all noted that their conceptualization of the MR option did differ across labels, sometimes affecting their MR endorsement. However, these respondents varied in their interpretation and conceptualization of the MR option across label variations. Upon each participant’s indication that MR option label did in fact influence their response selection, the interviewer asked a series of follow-up questions regarding their conceptualization of each label; these responses are discussed below.

Neutral. Although each of the three respondents mentioned that the *neutral* midpoint label differed from other common labels such as *neither agree nor disagree* and *undecided*, there was little agreement as to how they interpreted this label. One respondent suggested that he used *neutral* to indicate what Breckler (1994) would describe as ambivalence saying that his endorsement of response option indicates, “[I] really don't want to accept the strongly agree or disagree...it's like you're not picking sides, you're just kind of there...not necessarily on a number scale but you're just in-between...”. Conversely, another respondent seemed to be describing the expression of indifference when endorsing a MR option labeled *neutral*: “you really don’t care...”. Finally, the last respondent noted a difference between *neutral* and other MR option labels had an “impact” for these different labels, but failed to elaborate on any specific details regarding interpretation of *neutral* versus *neither agree nor disagree*.

Neither agree nor disagree. Again, respondents did not indicate a universal interpretation or conceptualization of the *neither agree nor disagree* MR option label, but they did note that it differed in meaning from other typical midpoint labels. One respondent suggested that *neither agree nor disagree* reflects a spot along a continuous “number scale” more so than an *undecided* midpoint label. This particular response suggests that this respondent may be more inclined to use a midpoint labeled *neither agree nor disagree* appropriately (i.e., as an indicator of a middling stance/level of a construct of interest) compared to a midpoint labeled *undecided*. Another respondent suggested that endorsement of a midpoint labeled *neither agree nor disagree* reflects respondent indecision or the fact that the respondent, “can’t take a stand” in response to an item. This respondent also noted a preference for selecting a MR option labeled

neutral versus *neither agree nor disagree* because of this interpretation. Finally, the third respondent indicated that a midpoint option labeled *neither agree nor disagree* would impact response selection more so than one labeled *neutral*, but did not elaborate on this with any specific detail.

Undecided. Although there was no definitive consensus on the interpretation of a MR option labeled *undecided* across respondents, based on responses it did seem to encourage MR option abuse as previously defined. That is, both respondents providing interpretation of a MR option labeled *undecided* noted that endorsement of such an option typically reflects something construct-irrelevant. For instance, one respondent noted a belief that respondents selecting *undecided* could, “have any other reason for it – it kind of felt like *undecided* was like an ‘other’...”. Perhaps similarly, another respondent mentioned that she had endorsed *undecided* when, “[I] don't think about it” or “[I] don't really have an answer”. Ultimately, both responses seemed to indicate that these respondents were more inclined to endorse a midpoint option labeled *undecided* for construct-irrelevant reasons as compared to the other MR option labels discussed.

No opinion. Based upon a number of the think-aloud participant responses and the sentiments they expressed while answering the structured interview questions, the interviewer felt inclined to ask about the distinction between midpoint endorsement and endorsement of an additional, *no opinion* response option. This decision was made when respondents seemed to be citing indifference as a motivator to endorse the midpoint. The interviewer proceeded to follow-up with respondents noting a difference across midpoint response option labels to clarify whether they distinguished between MR options and an additional *no opinion* response option. One respondent seemed to indicate a greater

feeling of indifference when selecting *no opinion* versus a MR option, mentioning that he would select *no opinion* after thinking about a stance/response to an item statement, but “just don’t care enough about the subject to give an answer”. Alternatively, another respondent stated that she may be more inclined to endorse an additional *no opinion* response option by default when she does not understand the item statement. However, this same respondent indicated that she would be more likely to endorse the MR option in this same situation if there was no additional *no opinion* response option offered, suggesting that some respondents may be inclined to endorse the midpoint when other more appropriate response options are not offered.

Study 2 Summary

A total of eight students participated in the think-aloud protocol and structured interview designed to address the two pre-specified research questions associated with Study 2. Given the extensive nature of the findings, overall results for Study 2 qualitative data analyses are briefly summarized in correspondence with each research question below.

Research Question 1: Do respondents appear to abuse the midpoint in practice? Do they provide construct-irrelevant justifications for endorsing the midpoint in response to the items administered? Yes, overall, it does appear that respondents abuse the MR option in practice, offering mostly construct-irrelevant justifications when endorsing the MR in response to the items administered during the think-aloud. More specifically, upon coding of MR endorsement justifications provided by think-aloud participants, the vast majority fell within the *Indecision* organizing theme category (21 out of 27 coded MR endorsement justifications). Of these 21 coded

statements falling within the *indecision* organizing theme, eleven were determined to reflect *non-commitment*, seven were determined to reflect *uncertainty*, and three were determined to reflect *indifference* (each of which represent a basic theme category subsumed within the *Indecision* organizing theme). Only six statements were determined to reflect one of the two remaining basic themes, either *ambivalence* (two count) or *item statement clarity* (four count) respectively.

Perhaps in further support of the propensity for respondents to abuse the MR in practice, there were some notable differences in frequency of MR endorsement across think-aloud items that seemed to coincide with certain item characteristics. For instance, the only SoI think-aloud item eliciting MR endorsement was item 4 (“I don’t know where I fit in the world”), which was the only reverse-scored item administered as part of the think-aloud. Additionally, during the think-aloud, MR was endorsed most frequently (four times) for two of the M-GUDS SELFUND subscale items (item 12, “Knowing about the experiences of people of different races increases my self-understanding”; item 4, “Knowing someone from a different ethnic group broadens my understanding of myself”). The content of both items deals with race or ethnicity, which may be more sensitive subject matter compared to the other think-aloud items. In this situation, participants may have felt uncomfortable providing their justification aloud for the researcher, opting instead to remain noncommittal and endorse the MR option.

Research Question 2: Do respondents conceptualize the midpoint response option differently depending on systematic variations in the midpoint response option label?

Yes, even though a slight majority of think-aloud participants indicated that MR option label made no impact on their item response selection (four participants out of seven), there were still three participants who indicated that MR option label did affect their conceptualization and endorsement of the MR. Although these three respondents indicated that differences in MR option label influenced their conceptualization of the MR, they did not demonstrate consistent interpretation of the MR across label variations, indicating that while variation in labels may affect MR conceptualization, it appears to do so differentially across respondents.

V. Discussion

The purpose of the current study was to build upon previous research examining the role of various item and respondent characteristics in MR abuse versus proper use in noncognitive assessment. Recall that MR abuse as it is defined within this dissertation is characterized by endorsement of the MR due to the influence of construct-irrelevant factors rather than a middling level of the construct of interest. When the MR is abused in such a way and construct-irrelevant variance is introduced within respondent scores, the validity of inferences made based upon noncognitive assessment scores is ultimately threatened. Once item and respondent characteristics potentially contributing to construct-irrelevant variance in scores are identified and the nature of their relationship to MR abuse is better understood, suggestions can be derived for best practice regarding use of the MR in assessment. That is, suggestions could be made for assessment practitioners regarding the decision to include or exclude the MR and/or attempts to control for various construct-irrelevant factors and their effects on MR abuse in practice.

To this end, a mixed-methods approach was utilized in attempts to better understand the various item and respondent characteristics that potentially contribute to MR abuse with the goal of informing and improving noncognitive assessment practice. Study 1, which was quantitative in nature, focused on expanding upon prior research by improving methodology to strengthen conclusions made based upon this line of research with the experimental manipulation of assessment characteristics. These experimental manipulations included the systematic variation of the following item characteristics: item set location (i.e., whether items were administered earlier versus later within an assessment instrument) and MR option label (i.e., holding the number of response scales

constant (1-7), but manipulating MR option label, either *unlabeled*, *neutral*, *undecided*, or *neither agree nor disagree*) for the same 20 noncognitive assessment items across randomly equivalent groups of respondents. A series of hierarchical generalized linear models were specified to examine the effect of these experimental manipulations, in addition to respondent characteristics (self-reported effort, verbal aptitude) (and their interactions) hypothesized to affect MR endorsement.

Study 2, which was qualitative in nature, sought to gain a more thorough understanding of the justifications respondents provide when endorsing the MR and their conceptualization of the MR upon manipulation of MR label (*unlabeled*, *neutral*, *undecided*, *neither agree nor disagree*). A think-aloud protocol was utilized to address the first research question associated with Study 2 in order for the researcher to gather unsolicited information regarding any construct-relevant or irrelevant justifications respondents may provide for MR endorsement. In addition, a series of structured interview questions were administered to think-aloud participants to address the second research question associated with Study 2. These follow-up questions specifically prompted participants to explain if and/or how MR option label influences their response selection.

Integration and Summary of Findings from Study 1 and Study 2

Given the results of both studies, there is reason to believe that the MR option is subject to abuse by respondents in noncognitive assessment. Upon administration of a series of noncognitive assessment items developed to measure a certain construct of interest, it does not appear as though respondent MR endorsement solely reflects middling levels of this construct. Instead, MR endorsement appears to be affected in

nuanced ways by a myriad of construct-irrelevant factors, both measurement- and respondent-specific.

Results of Study 1 Model 6 suggest that each of the respondent and experimentally manipulated item characteristics initially thought to influence probability of MR endorsement do so in some manner. Furthermore, for the most part (excluding the *neutral* and *undecided* MR option conditions), the effects of item set location, respondent levels of effort, and levels of verbal aptitude (although to a lesser extent) appear to affect MR endorsement in ways that were previously hypothesized. For the *unlabeled* and *neither agree nor disagree* MR option conditions, each of these characteristics affected probability of MR endorsement in such a way that would be consistent with Krosnick's (1999) theory of satisficing in which respondents exerting lower levels of cognitive effort are more likely to satisfice (sometimes in the form of MR endorsement rather than a more appropriate response option). That is, as item set location increased (items appeared later in the instrument), respondent levels of effort decreased, and as levels of verbal aptitude (SAT verbal scores) decreased, probability of MR endorsement increased. The significant effect of item set location was particularly notable as there were only 27 additional items appearing before the SoI and M-GUDS items in the later versus earlier item set location conditions.

As mentioned above, the relationships amongst the respondent characteristics and experimentally manipulated item characteristics in Model 6 were less intuitive for the *neutral* and *undecided* MR option conditions. For the *neutral* condition, the relationship between effort and MR endorsement depended on item set location and SAT verbal score of the respondent. The significance of this interaction appeared to be driven by the

difference in MR endorsement across respondents with high versus low SAT verbal scores reporting lower levels of effort. In this situation, when respondents had higher SAT verbal scores, probability of MR endorsement was much higher than when respondents had lower SAT verbal scores. For the *undecided* condition, MR label and SAT verbal interacted in such a way that respondents with high SAT verbal scores were more likely to endorse the MR option than those with low SAT verbal scores. Neither of these effects is consistent with what was originally hypothesized based upon Krosnick's (1999) theory of satisficing, but the fact that both of these MR labels interact with other variables to influence probability of MR endorsement differently than in the unlabeled condition provides evidence to support the hypothesis that various MR labels function differently in practice.

Interestingly, findings from both studies suggest that different MR options do not necessarily function in the same way across respondents. At least for some respondents, it seems that variation in MR option label interacts with other variables (related to an assessment item/instrument, the respondent, or to both) to influence probability of MR endorsement in an idiosyncratic manner. With regard to Study 1, respondent characteristics such as self-reported effort and verbal aptitude affected endorsement of the MR in similar, more predictable ways for the *unlabeled* and *neither agree nor disagree* conditions, whereas the nature of these interactions were rather unique and slightly less predictable for *neutral* and *undecided* (as evidenced by their respective four-way and two-way significant interaction effects in Model 6). With regard to Study 2, four out of seven respondents indicated that variation of MR option label in the think-aloud made no difference in their endorsement of the MR option. In direct contrast, there were

three other respondents stating that changes in MR option label did influence their conceptualization and endorsement of the MR option (though not in consistent ways).

Implications of Findings for Assessment Practice

Taken into consideration together, findings from both studies seem to suggest that it is likely very difficult to foresee all of the ways in which and to what extent various construct-irrelevant factors affect MR abuse in practice. However, what we do know is that there are certain item and respondent characteristics included within the current study that seem to affect probability of MR endorsement in some notable ways. These variables introduce construct-irrelevant variance within respondent scores and threaten the validity of the inferences practitioners may wish to make based upon these scores. As such, findings from Studies 1 and 2 have important implications for researchers and practitioners alike who administer instruments with Likert-type response scales.

Researchers or practitioners planning to administer self-report assessment instruments may consider avoiding the issue of MR abuse by excluding the MR option altogether. Findings from both studies indicate that there appears to be no MR option (*unlabeled, neutral, undecided, neither agree nor disagree*) that is not prone to abuse. Although some appear to be affected in more predictable ways than others (such as *unlabeled* and *neither agree nor disagree*), no one MR option is treated alike across all respondents. Accordingly, no definitive recommendation can be made with regard to best practice in selections of MR option label. This leaves researchers and practitioners looking to utilize noncognitive assessments with an important decision – whether to include or exclude the MR option. When deciding to administer an assessment instrument, researchers have the option to modify response scale options however they

see fit. This includes the decision to include or exclude the midpoint amongst response scale options. If they wish to exclude the MR option, they can simply offer an even number of response scale options. However, they must do so realizing that these alterations, or customizations, have the potential to influence response selection and in turn affect the functioning and psychometric properties of the scale. As noted by Presser and Schuman (1980), the decision to exclude the MR may have negative implications, as respondents intending to use the MR properly would be forced to select another, less appropriate response option amongst those provided, resulting in the introduction of error within respondent scores.

The obvious alternative to not offering the MR option would be to offer it. If practitioners wish to offer the MR option, they can attempt to control for, or minimize, the effects of construct-irrelevant factors influencing MR abuse. In this is the case, there are several steps that researchers or practitioners may consider taking in the future based upon the results of the current study.

First, practitioners should make attempts to bolster respondent levels of effort expended on noncognitive assessments. The significant main effect for effort in Study 1 Model 6 suggests that respondent levels of self-reported effort affect probability of MR endorsement, regardless of item set location and SAT verbal score, across three of the four MR option label conditions - *unlabeled*, *undecided*, and *neither agree nor disagree*. Across all three of these conditions, probability of MR endorsement increased as levels of self-reported effort decreased. In the *neutral* MR option label condition, effort also influenced MR endorsement, but the relationship between effort and midpoint endorsement was dependent upon item set location and the SAT verbal score of the

respondent. Although the nature of the relationship between effort and midpoint endorsement differed for the *neutral* condition, it is evident that effort impacts probability of MR endorsement in some manner regardless of midpoint label. Moreover, levels of respondent effort may also be related to feelings of indifference cited as justification for MR endorsement during the think-aloud protocol in Study 2. If respondent levels of effort were increased, perhaps respondents would be less likely to feel indifferent. In order to increase respondent levels of motivation, practitioners may consider implementing a number of interventions such as raising the stakes of assessment testing, offering incentives for better performance, or attempting to inform students of the importance of performing well on low-stakes assessment.

Second, practitioners should strive to ensure that assessment item statements are clear, concise, understandable, and relevant to the target respondent population. This suggestion is made largely based upon the findings of Study 2, where think-aloud participants repeatedly justified endorsement of the MR because they were confused, uncertain, or indifferent with regards to what an item statement was asking. Additionally, although the main effect of SAT verbal was non-significant in Model 6, probability of MR endorsement was consistently higher for respondents with lower SAT verbal scores across most conditions in Study 1 (excluding the *neutral* and *undecided* MR label conditions due to significant two-way interaction between the *undecided* midpoint label condition and SAT verbal and the four-way interaction amongst the *neutral* midpoint label, item set location, effort, and SAT verbal).

There is some additional evidence from both studies that may further support this particular suggestion. In both studies, SoI item 3 (“I have a clear set of personal values or

moral standards”) and M-GUDS SELFUND item 4 (“Knowing someone from a different ethnic group broadens my understanding of myself”) were among the least and most likely items, respectively, to elicit MR endorsement from respondents. According to simple descriptive statistics reported in Study 1, SoI item 3 was associated with the lowest average incidence of MR endorsement whereas M-GUDS item 4 was associated with the highest average instance of MR endorsement. Consistently, in Study 2, no respondent endorsed the MR in response to SoI item 3, whereas four respondents (tied for the most respondents endorsing the MR in response to an item administered during the think-aloud) endorsed the MR in response to M-GUDS item 4. Given the nature of these two item statements, their respective item content, clarity, and relevance could have played a role in MR endorsement in both instances. Accordingly, researchers and practitioners may consider conducting think-alouds with a sample of respondents from the target population prior to data collection. Conducting think-alouds with a sampling of respondents from the target population may not only help practitioners to ensure that items are clear, concise, understandable, and relevant, but also reveal if any other item-specific characteristics are systematically influencing response selection (i.e., if items are extremely divisive, or if content is especially sensitive for respondents, limiting variance in response selection) .

Third, practitioners should attempt to minimize the length of assessment instruments. Given the results of Study 1, probability of MR endorsement tends to increase when an identical item is administered later versus earlier in an assessment instrument. Although the main effect of item set location was no longer statistically significant in Model 6 as it was in Models 2 through 5, its positive value still indicates

that items administered later tend to result in increased midpoint endorsement across most conditions. If limiting the length of an assessment is not an option, the practitioner may consider enacting an intervention of some kind in attempts to bolster respondent engagement throughout the entirety of the assessment, perhaps by offering breaks and/or re-orienting respondents to the task at hand.

Finally, practitioners should consider offering an additional response option such as *no opinion*, *don't know*, or *not applicable* as an alternative to the MR option. Given some of the responses provided by think-aloud participants in their justifications for MR endorsement and in response to structured interview questions, it appeared that some respondents would benefit from having a different, perhaps more appropriate, response option in addition to the MR. During the think-aloud, there were numerous respondents who cited feelings of uncertainty or indifference as justification for MR endorsement. Additionally, the three think-aloud participants indicating that MR option label affected their MR selection also appeared to convey this sentiment in response to structured interview questions. Perhaps if these respondents were provided with a more appropriate response option that allows them to express their feelings of indifference or uncertainty, the MR option would be less prone to abuse. Research conducted by Harter (1997) and Bishop et al. (1988) lend additional support this conclusion. Results of each study indicate that respondents discriminate between available MR options and additional nonresponse options such as *don't know* or *no opinion*. Furthermore, Harter found that offering an additional *don't know* response option actually appeared to improve the functioning of the MR option (more closely reflecting a middling level of the construct of interest).

Limitations of the Current Study and Directions for Future Research

Findings from the current study could be confounded by a number of limitations. Most of these potential confounds stem from the data collection context which was a university-wide large-scale assessment testing session held for institutional accountability purposes. This context has several unique features that can possibly influence the nature of and limit the generalizability of the findings from the current study. First, this university-wide testing session is low-stakes for students. Although students are required to attend these assessment testing sessions by the university, they have no direct consequences for poor performance, thus there is no external motivation for them to expend effort on these tests. Certain provisions are made during this assessment session in attempts to bolster student levels of motivation including the formal training/presence of test session proctors within each room and standardized test session protocol/instructions for students that emphasize the importance of these assessments to the university; however, effort expended by students remains a concern in this context as the test session is lengthy (approximately 3 hours), some tests are cognitively taxing, and there are no consequences for individual students based upon performance. As such, the results of the current study may not generalize to a high-stakes assessment context. Second, data used were collected during the fall university-wide assessment testing session, thus the sample for the current study was restricted to incoming first-year students. Given this restriction, the sample was rather homogeneous in terms of age, ethnicity, sex, and cognitive ability, thus posing a threat to the generalizability of the findings to other populations. Third, the assessments completed by participants in Study 1 (SoI and M-GUDS, 20 items total) were not administered in isolation. As mentioned

previously, students complete a series of cognitive and noncognitive assessment tests over the duration of the 3 hour university-wide assessment testing session, with some respondents completing the SDA-7 third and others fourth in the test configuration. Accordingly, there could be any number of factors from previous assessments, or the nature of the assessment session itself, that contribute to the nature of the findings within the current study.

An additional potential confound within the current study is the measurement of effort. More specifically, the effort score utilized within the current study was calculated based upon student self-report data in response to the five Effort subscale items administered on the 10-item SOS instrument. In addition to the SOS being a self-report measure, student responses did not correspond to the SDA-7 on which the SoI and M-GUDS items were administered. The SOS was the last instrument administered to students during the assessment testing session in order to gauge overall levels of respondent effort expended on the assessments they completed throughout the session. As such, respondents were instructed to respond to SOS items accordingly, thus their effort scores may not reflect those they put forth on the SDA-7. Moreover, SOS item response scale options include a midpoint response option labeled *neutral*. If the MR option is prone to abuse in practice as the current study suggests it is (and this effect carries over to the SOS), we ultimately cannot draw valid conclusions based upon respondent SOS scores. It would be better to avoid using self-report measures of effort that use Likert-type response scales in future studies. Instead, researchers may consider using alternative indicators of respondent effort such as response time information or conducting a study using two groups of respondents with differing levels of motivation (one group

participating in low-stakes assessment versus another participating in high-stakes assessment).

Finally, the nature of the constructs the SoI and M-GUDS items were developed to measure may be a potential confound within the current study. That is, the SoI and M-GUDS items administered within the current study were limited in scope to sense of identity and ease and comfort with diversity. Due to the nature of these constructs, responses may require greater levels of maturity or more meta-cognitive skills on the behalf of respondents. Furthermore, item content may be more sensitive or item statements could be more complex than those from other noncognitive assessment instruments, all potentially eliciting greater MR endorsement. The current study should be replicated using alternative instruments.

Future studies should seek to replicate and expand upon the findings from the current study. With regard to the limitations described above, future studies examining the nature of MR abuse in assessment practice may be strengthened by using an alternative measure of effort and noncognitive assessment instruments that are less sensitive in nature and include item statements that are easier for respondents to read and understand. Additionally, more research could be conducted using data already collected for the purpose of the current study. These additional analyses could focus on examining the psychometric properties (i.e., reliability and validity) of the SoI and M-GUDS scores across the experimental conditions. To this end, these studies could focus on examining any differences in reliability estimates, relationships amongst SoI and M-GUDS scores and external variables, and factor structure of the SoI and M-GUDS across variations in experimentally manipulated factors such as item set location and MR option label.

Findings from Study 1 and 2 suggest that there may be additional item characteristics influencing probability of MR endorsement currently unaccounted for within Study 1 models. Thus, future studies could further examine the nature of different item characteristics possibly influencing probability of MR endorsement. Possible additional item characteristic predictors include: an indicator of item statement clarity or readability (e.g., readability index score), an indicator of item statement sensitivity (e.g., dummy-coded variable to indicate more sensitive item content, regarding personal morals, feelings about other ethnicities), an indicator of more strongly worded item statements (e.g., dummy-coded variable to indicate more divisive item content), or an indicator of item reverse-scoring (e.g., dummy-coded variable identifying reverse-scored items).

Finally, potential follow-up studies may seek to better understand the more complex, counterintuitive effects noted in Model 6 for the *neutral* and *undecided* MR option conditions. These could include more in-depth examination of these relationships across specific groups of respondents. More specifically, for the *neutral* condition, the researcher may look more closely at the nature of the significant four-way interaction effect that appears to be driven by respondents reporting lower levels of effort, receiving items earlier in the assessment instrument, with higher SAT verbal scores (e.g., how many of these respondents exist within the data set, does this effect appear to be driven by extreme outliers?, if there is a substantial number of respondents in this group, is there research literature to explain or support their behavior?). The researcher may consider completing another similar investigation for the *undecided* condition in which

respondents with higher SAT verbal scores exhibited higher probabilities of MR endorsement.

General Conclusion

In summary, findings from both studies suggest that the MR option is prone to abuse in practice. More specifically, there is reason to believe that various item and respondent characteristics (as well as their interactions) have the potential to significantly affect MR option endorsement. When interpreting Model 6 with respect to the research questions associated with Study 1, we see that item set location, MR option label, respondent self-reported levels of effort expended on assessments, and respondent levels of verbal aptitude all seem to affect MR endorsement in some shape or form – sometimes via complex interactions. In summarizing the findings of Study 2, we see that justifications provided by respondents for MR endorsement are mostly construct-irrelevant and the effects of variation in MR option label are largely dependent on the individual respondent.

If respondent endorsement of the MR option is indeed affected by item characteristics such as item set location and MR option label, as well as respondent characteristics such as effort and SAT verbal score, this supports the claim that measurement and respondent dispositional characteristics simultaneously affect endorsement of a MR. The results of the current study should be replicated before any definitive conclusions can be made regarding the nature of MR abuse in noncognitive assessment practice. However, preliminary findings do seem to suggest that certain item and respondent characteristics influence MR. Thus, future study is necessary in order to

better understand the conditions and characteristics that contribute to MR abuse in order to limit their potential biasing effects on respondent scores.

Footnotes

ⁱDue to the non-linear nature of the relationship between the logit and probability, the probability of midpoint response endorsement will equal the median probability instead of the mean probability in the population. However, it is possible to employ a population-average model in order to obtain an estimate of γ_{00} that when converted to a probability will more closely resemble the sample mean probability of endorsing the midpoint response option and equal the mean probability of endorsing the midpoint response option in the respondent population. Typically, the choice to employ a population-average model that yields such an estimate rather than a unit-specific model which does not is a function of the kind of inferences a researcher wishes to make based upon their data. Such inferences include those that extend to the relationship amongst various item and respondent characteristics that influence probability of an individual respondent endorsing the midpoint response option (unit-specific) versus those that extend to the relationship amongst these same item and respondent characteristics and the average probability of midpoint response endorsement in a group of respondents (population-average) (Raudenbush & Bryk, 2002).

For the current study, the decision to employ a population-average model was made based upon statistical concerns. More specifically, within the current study, probability of midpoint response option endorsement is likely to be positively skewed, increasing the likelihood of violating the assumption that random effects included within Equation 3 are normally distributed. In order to statistically control for this, the population-average model was utilized as it has historically proven to be more robust to the violation of such assumptions (Raudenbush & Bryk, 2002).

Appendix A

Item content and scoring of the Revised Sense of Identity Scale (SoI)

Administered in Study 1

****SoI-1 (no midpoint label; reverse-scored item is 4)****

The following set of questions deals with how you feel about yourself. Please bubble in the number from **1 to 7** that indicates the extent to which you agree or disagree with that statement.

Please take your time and answer thoughtfully.

Strongly Disagree						Strongly Agree
1	2	3	4	5	6	7

1. I have a definite sense of purpose in life.
2. I know what I want out of life.
3. I have a clear set of personal values or moral standards.
4. I don't know where I fit in the world.*
5. I have specific personal goals for the future.
6. I have a clear sense of who I want to be when I am an adult.

Appendix B

Item content and scoring of the Revised Miville-Guzman University-Diversity Scale

(M-GUDS)

Administered in Study 1

****MGUDS-1 (no midpoint label; reverse-scored items are 1, 3, 6, 8, 9, 13)****

The following statements relate to how you interact with others. Using the scale below, please indicate the extent to which you agree with each statement by selecting a number from 1 through 7. *There are no right or wrong answers, just respond as accurately as possible.*

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

1. It's hard to find things in common with people who have different sexual orientations.* (ECSO)
2. In getting to know someone, I like knowing both how he/she differs from me and is similar to me. (SIMDIF)
3. I often feel irritated by persons of a different race.* (DRD)
4. Knowing someone from a different ethnic group broadens my understanding of myself. (SELFUND)
5. Knowing about the different experiences of other people helps me understand my own problems better. (SELFUND)
6. It's hard for me to feel close to a person who has a different sexual orientation.* (ECSO)
7. Knowing how a person differs from me greatly enhances our friendship. (SIMDIF)
8. It's really hard for me to feel close to a person from another race.* (DRD)

9. I am only at ease with people of my own race.* (DRD)
10. I feel comfortable getting to know someone who has a different sexual orientation. (ECSO)
11. I can best understand someone after I get to know how he/she is both similar and different from me. (SIMDIF)
12. Knowing about the experiences of people of different races increases my self-understanding. (SELFUND)
13. Getting to know someone of another race is generally an uncomfortable experience for me.* (DRD)
14. In getting to know someone, I try to find out how I am like that person as much as how that person is like me. (SIMDIF)

Appendix C

General Instructions for Study 2 Participants

General Introduction/Instructions for Golden Ticket Holders

Welcome, and thank you for coming. I'm Chris and this is Becca, and we're PhD students in Assessment & Measurement here at JMU. We're going to start by giving you a quick overview of what we'll be doing here this (morning/afternoon). You can ask any questions you might have, and then you can decide if you want to stay here and participate in the research OR go back to 2301 and do the standard assessment testing.

First, I'm sure you're wondering about time. Participation in our research should take about 2 hours, which is probably 30 minutes less than the amount of time you'd spend in 2301. Basically, we want to know what you think about several of our tests and questionnaires. It's part of what they call the validation process—to make sure tests are measuring what they were designed to measure. Part of the validation process entails identifying if there are any aspects of the tests (such as instructions, items, etc.), that are problematic. You all have been randomly selected to help us with that, if you choose to do so.

If you choose to participate, for most of the session, you'll be working independently in this room. We'll give you a packet of tests and questionnaires and ask you to take them as you normally would—but also providing some written feedback along the way. In addition, we're conducting some one-on-one research using a think-aloud design. Because that involves talking, we'll have to do those short sessions in nearby rooms. The way we'll organize it is like this: We'll get you started on the packet, then one person will go with X and one with me for 20-30 minutes. When the session is over, you'll come back to this room and continue working on the packet; then the next person will go to do a think-aloud, and so on. When everybody has finished the think-alouds and the packets, we'll dismiss you all.

Do you have any questions?

OK, now it's time to decide what you want to do. If you'd like to participate in the research, then please sign the consent form on your desk. If you'd rather go back to the standard assessment testing, you can grab your things and return to 2301.

Begin data collection.

Appendix D

Study 2 Consent Form

Consent to Participate in Assessment Day Research

Purpose of Study

You are being asked to participate in research conducted by the Center for Assessment and Research Studies (CARS) at James Madison University. The purpose of this study is to gather verbal and written feedback from JMU students about several assessment tests and questionnaires.

Potential Risks & Benefits

The researchers do not perceive more than minimal risks from your involvement in this study. You may feel uncomfortable about some of the questions on the questionnaires. If this is the case, you may choose not to respond to these questions. However, all the information you provide will be kept completely confidential and viewed/heard only by CARS researchers. The feedback you provide will be de-identified so that it is not linked to your name or ID number. Potential benefits from participation in this study include contributing to the validity and effectiveness of assessment testing.

Research Procedures

Should you decide to participate in this research study, you will be asked to sign this consent form once all your questions have been answered to your satisfaction. You will be asked to complete a packet of tests and questionnaires, providing written feedback along the way. In addition, you will be interviewed one-on-one by a researcher using a think-aloud protocol. Your responses will be audio-recorded so that they can later be analyzed for qualitative themes.

Time Required

Participation in this study will require approximately 120 minutes of your time.

Participation

Your participation is entirely voluntary. You are free to participate in standard assessment testing activities instead. Either option will fulfill your assessment day obligations.

Giving of Consent

I have read this consent form and I understand what is being requested of me as a participant in this study. I freely consent to participate. I have been given satisfactory answers to my questions.

I give consent to be audio-recorded during my think-aloud. _____ (initials)

Name of Participant (Printed)

Participant's JMU ID number

Name of Participant (Signed)

Date

Appendix E

Study 2 Pilot Testing Student Feedback Instructions

Fall 2012 Assessment Day Pilot Testing Student Feedback

Thank you participating in this study! We greatly appreciate your time and effort. As we mentioned previously, you will be working independently to complete a packet of tests and questionnaires in this room until you are called to participate in the alternate research activity. Please complete each test and questionnaire as you would during a normal testing session – reading directions for each test/questionnaire, reading each question or item statement, and then responding accordingly by recording your answer on the test form (you may write on the test). In addition to completing these tests and questionnaires as you normally would, we request that you provide additional feedback for each item following the directions provided to you on this document. This feedback will provide us with additional information that will help to improve the quality of the assessments administered on Assessment Day.

You will be completing a couple tests and/or questionnaires. The proctor will administer the first instrument to all students at the beginning of the study. Please complete the first instrument by reading the directions, responding to each question/item statement on the test form. In addition, please follow the directions provided here to provide additional information for each item you complete.

Please provide additional information for each item statement that you complete according to the table below. For example, if you like an item, please place a “+” directly next to your corresponding item response on the test form, if you dislike an item, place a “-“ directly next to your corresponding item response on the test form, and if you are confused by an item, please place a “?” directly next to your corresponding item response on the test form. Finally, please provide any comments or clarification next to the +, -, or ? you have placed next to each item response to provide details regarding your opinions of the item. See the example below for clarification.

Once you have completed the first instrument, you may request the next instrument and receive instructions from the proctor regarding how to proceed. If you should have any questions or concerns while completing these instruments, please ask the proctor. They will be happy to answer any of your questions and provide clarification – they should be seen as a resource.

Example:

Item statement	Response	+, -, ?	Comments
1	7	+	I really liked this item. It was easy to understand and I think it does a good job of measuring how motivated I am to do well in my classes this semester.
2	4	-	I don't really like this item. I feel like I could answer this item differently depending on what the researcher intends.
3	4	?	I thought this item was confusing. It was too long and wordy. I didn't know how to answer it.

Appendix E

Study 2 Think-aloud Protocol Instructions

Researcher Introduction/Practice/Instructions

I'm going to ask you to complete a short questionnaire. As you are responding to each statement, I'd like you to think aloud—that is, say all the things that go through your mind as you're choosing your answer. I'll demonstrate that process using the first statement on this practice questionnaire.

OK, I see that the response scale goes from 1 to 7, with these choices (SD > SA). The first statement says, "I'm an impulsive person." Hmmmm...what does that mean? That I do or say things without thinking about them first. Well, I am impulsive most of the time. Not always, but a lot...so I would say 6 - "Agree".

That example was just one way people might think through their answers. There's no "right" way to do it. I just want you to say all the things that go through your mind as you're choosing your answers. Does the thinking-aloud thing make sense? OK, you go ahead and try the next couple.

Give feedback/clarification/encouragement as needed for items 2 and 3. If additional practice is needed, item 4 can be used.

Now I'm going to **turn on the recorder** and we'll move on to the real questionnaire. Remember to talk out loud—say everything you're thinking as you figure out your answer choices. You don't have to write down your answers, because they're not the focus of the study. I'm more interested in the process by which you reach your answers. Does that make sense?

You will notice that some of the item statements are fairly similar across the different scales. That's intentional; I'm researching different scales that measure similar traits, so the item wording will be similar across the scales. Be sure to read the instructions and note the response options for each of the scales. Then think aloud as we did in the practice when you respond to each item statement. Do you have any questions before we start? Alright, let's begin.

During process, can prompt by saying, "Remember to tell me what you're thinking."

Appendix G

Study 2 Think-aloud Practice Questionnaire

Practice Questionnaire

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

1. I'm an impulsive person.
2. I tend to remember important meetings and appointments.
3. I'm an organized person.
4. I'm usually able to pay attention during conversations.

Appendix H

Study 2 Think-aloud Structure Interview Questions

Structured Interview Questions

Great, thank you! We're almost done, but before you go back to the other room I want to ask you a couple of quick questions. And I'll keep the recorder on if that's OK, so I don't forget your answers.

First, did you notice a difference in response options across the scales? If so, what was it?

Second, did this make a difference in how you responded to the item statements across the scales? How so?

Thank you. Do you have any questions about anything?

Turn off recorder.

*

Appendix I

Study 2 Think-aloud Questionnaire Forms A-H

Think-Aloud Questionnaire – Form A

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

1. I have a definite sense of purpose in life.
2. Knowing about the experiences of people of different races increases my self-understanding.
3. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

4. I know what I want out of life.
5. Knowing someone from a different ethnic group broadens my understanding of myself.
6. I can best understand someone after I get to know how he/she is both similar and different from me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

7. I have a clear set of personal values or moral standards.
8. Knowing about the different experiences of other people helps me understand my own problems better.
9. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

10. I don't know where I fit in the world.
11. I have specific personal goals for the future.
12. Knowing how a person differs from me greatly enhances our friendship.

Think-Aloud Questionnaire – Form B

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

1. I don't know where I fit in the world.
2. I have specific personal goals for the future.
3. Knowing how a person differs from me greatly enhances our friendship.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

4. I have a definite sense of purpose in life.
5. Knowing about the experiences of people of different races increases my self-understanding.
6. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

7. I know what I want out of life.

8. Knowing someone from a different ethnic group broadens my understanding of myself.

9. I can best understand someone after I get to know how he/she is both similar and different from me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

10. I have a clear set of personal values or moral standards.

11. Knowing about the different experiences of other people helps me understand my own problems better.

12. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Think-Aloud Questionnaire – Form C

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

1. I have a clear set of personal values or moral standards.
2. Knowing about the different experiences of other people helps me understand my own problems better.
3. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

4. I don't know where I fit in the world.
5. I have specific personal goals for the future.
6. Knowing how a person differs from me greatly enhances our friendship.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

7. I have a definite sense of purpose in life.
8. Knowing about the experiences of people of different races increases my self-understanding.
9. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

10. I know what I want out of life.
11. Knowing someone from a different ethnic group broadens my understanding of myself.
12. I can best understand someone after I get to know how he/she is both similar and different from me.

Think-Aloud Questionnaire – Form D

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

1. I know what I want out of life.
2. Knowing someone from a different ethnic group broadens my understanding of myself.
3. I can best understand someone after I get to know how he/she is both similar and different from me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

4. I have a clear set of personal values or moral standards.
5. Knowing about the different experiences of other people helps me understand my own problems better.
6. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

7. I don't know where I fit in the world.
8. I have specific personal goals for the future.
9. Knowing how a person differs from me greatly enhances our friendship.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

10. I have a definite sense of purpose in life.
11. Knowing about the experiences of people of different races increases my self-understanding.
12. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Think-Aloud Questionnaire – Form E

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

1. I don't know where I fit in the world.
2. I have specific personal goals for the future.
3. Knowing how a person differs from me greatly enhances our friendship.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

4. I have a clear set of personal values or moral standards.
5. Knowing about the different experiences of other people helps me understand my own problems better.
6. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

7. I know what I want out of life.
8. Knowing someone from a different ethnic group broadens my understanding of myself.
9. I can best understand someone after I get to know how he/she is both similar and different from me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

10. I have a definite sense of purpose in life.
11. Knowing about the experiences of people of different races increases my self-understanding.
12. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Think-Aloud Questionnaire – Form F

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

1. I have a definite sense of purpose in life.
2. Knowing about the experiences of people of different races increases my self-understanding.
3. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

4. I don't know where I fit in the world.
5. I have specific personal goals for the future.
6. Knowing how a person differs from me greatly enhances our friendship.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

- 7. I have a clear set of personal values or moral standards.
- 8. Knowing about the different experiences of other people helps me understand my own problems better.
- 9. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

- 10. I know what I want out of life.
- 11. Knowing someone from a different ethnic group broadens my understanding of myself.
- 12. I can best understand someone after I get to know how he/she is both similar and different from me.

Think-Aloud Questionnaire – Form G

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

1. I know what I want out of life.
2. Knowing someone from a different ethnic group broadens my understanding of myself.
3. I can best understand someone after I get to know how he/she is both similar and different from me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

4. I have a definite sense of purpose in life.
5. Knowing about the experiences of people of different races increases my self-understanding.
6. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

7. I don't know where I fit in the world.
8. I have specific personal goals for the future.
9. Knowing how a person differs from me greatly enhances our friendship.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

10. I have a clear set of personal values or moral standards.
11. Knowing about the different experiences of other people helps me understand my own problems better.
12. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Think-Aloud Questionnaire – Form H

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree

1. I have a clear set of personal values or moral standards.
2. Knowing about the different experiences of other people helps me understand my own problems better.
3. In getting to know someone, I try to find out how I am like that person as much as how that person is like me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

4. I know what I want out of life.
5. Knowing someone from a different ethnic group broadens my understanding of myself.
6. I can best understand someone after I get to know how he/she is both similar and different from me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree						Strongly Agree

7. I have a definite sense of purpose in life.
8. Knowing about the experiences of people of different races increases my self-understanding.
9. In getting to know someone, I like knowing both how he/she differs from me and is similar to me.

Please respond by indicating how much you agree or disagree with each statement using the response options **1** (Strongly Disagree) to **7** (Strongly Agree).

There are no right or wrong answers; just answer thoughtfully and honestly.

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree

10. I don't know where I fit in the world.
11. I have specific personal goals for the future.
12. Knowing how a person differs from me greatly enhances our friendship.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Attride-Stirling, J. (2001). Thematic networks: An analytical tool for qualitative research. *Qualitative Research, 1*, 385-405.
- Baumgartner, H., & Steenkamp, J. E. M., (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143-156.
- Bishop, G. F., Hippler, H., J., Schwarz, N., & Strack, F. (1988). A comparison of response effects in self-administered surveys. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 321-340). New York: John Wiley & Sons.
- Bogart, L. (1967). No opinion, don't know, and maybe no answer. *Public Opinion Quarterly, 31*, 332-345.
- Bollen, K. A., & Hoyle, R. H. (1990). Perceived Cohesion: A Conceptual and Empirical Examination. *Social Forces, 69*, 479 – 504.
- Breckler, S. J. (1994). A comparison of numerical indexes for measuring attitude ambivalence. *Educational and Psychological Measurement, 54*, 350-365.
- Campbell, Bonacci, Shelton, Exline, & Bushman (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment, 83*, 29–45.

- Coombs, C. H., & Coombs, L. C. (1976). "Don't know": Item ambiguity or respondent uncertainty? *Public Opinion Quarterly*, 40, 497-514.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475-494.
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3-31.
- DeMars, C. E., & Erwin, T. D. (2004). Scoring *neutral or unsure* on an identity development instrument for higher education. *Research in Higher Education*, 45, 83-95.
- DeMars, C. E., & Erwin, T.D. (2005, August). *Neutral or unsure: Is there a difference?* Poster presented at the annual meeting of the American Psychological Association, Washington, D.C.
- Edwards, A. L. (1946). A critique of "neutral" items in attitude scales constructed by the method of equal appearing intervals. *Psychological Review*, 53, 159-169.
- Elliot, A., & McGregor, H. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501-519.
- Erwin, T. D. (1991). *Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes*. San Francisco: Jossey-Bass.
- Ewell, P. T. (2002). An emerging scholarship: A brief history of assessment. In T.W. Banta & Associates. *Building a scholarship of assessment*. (pp. 3-25). San Francisco: Jossey-Bass.

- Finney, S. J., Pieper, S. L., & Barron, K. E. (2004). Examining the psychometric properties of the achievement goal questionnaire in a general academic context. *Educational and Psychological Measurement, 64*, 365–382.
- Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin, 69*, 192-203.
- Harter, J. K. (1997). *The psychometric utility of the midpoint on a Likert scale*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses. (UMI No. AAI9730273)
- Harzing, A. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management, 6*, 243-266.
- Kalton, G., Roberts, J., & Holt, D. (1980). The effects of offering a middle response option with opinion questions. *The Statistician, 29*, 65–78.
- Kieruj, N. D., & Moors, G. (2010). Variations in response style behavior by response scale format in attitude research. *International Journal of Public Opinion Research, 22*, 320-342.
- Klopfers, F. J., & Madden, T. M. (1980). The middlemost choice on attitude items: Ambivalence, neutrality, or uncertainty? *Personality and Social Psychology Bulletin, 6*, 97-101.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.
- Kyllonen, P. C. (2005). *ETS R&D Connections: The case for noncognitive assessments*. Retrieved from http://www.ets.org/Media/Research/pdf/RD_Connections3.pdf.

- Levine-Brown, P., Bonham, B. S., Saxon, D. P., & Boylan, H. R. (2008). Affective assessment for developmental students, part 2. *Research in Developmental Education, 22*, 1-4.
- Lounsbury, J. W., Levy, J. J., Leong, F. T., & Gibson, L. W. (2007). Identity and personality: The Big Five and narrow personality traits in relation to sense of identity. *Identity: An International Journal of Theory and Research, 7*, 51-70.
- Lounsbury, J. W., Huffstetler, B. C., Leong, F. T., & Gibson, L. W. (2005). Sense of identity and collegiate academic achievement. *Journal of College Student Development, 46*, 501-514.
- Marsh, K. R., & Pastor, D. A. (2011, October). *Apathy or indecision: An examination of characteristics that influence endorsement of a midpoint response*. Paper presented at the annual meeting of the Northeastern Educational Research Association, Hartford, CT.
- Marsh, K. R., & Pastor, D. A. (2012). Item and student characteristics that influence endorsement of a midpoint response option: A validity issue. Unpublished manuscript.
- Miville, M. L., Gelso, C. J., Pannu, R., Lui, W., Touradji, P., Holloway, P., & Fuentes, J. (1999). Appreciating similarities and valuing differences: The Miville-Guzman Universality-Diversity Scale. *Journal of Counseling Psychology, 46*, 291-307.
- Moely, B. E., McFarland, M., Miron, D., Mercer, S., & Illustre, V. (2002). Changes in college students' attitudes and intentions for civic involvement as a function of service-learning experiences. *Michigan Journal of Community Service Learning, 9*, 18-26.

- Nunnally, J. C. (1967). *Psychometric Theory*. New York: McGraw Hill.
- O'Muircheartaigh, C., Krosnick, J. A., & Helic, A. (2000). Middle alternatives, acquiescence, and the quality of questionnaire data. Retrieved, October 1, 2009, from <http://harrisschool.uchicago.edu/About/publications>.
- Pastor, D. A., & Cotten, M. (2009). *An exploratory factor analysis of the Miville-Guzman University-Diversity Scale*. Unpublished manuscript.
- Payne, Stanley L. (1951). *The Art of Asking Questions*. Princeton: Princeton University Press.
- Pieper, S. (2003). Refining and extending the 2 x 2 achievement goal framework: Another look at work-avoidance. *Dissertation Abstracts International*, 64, 4436A.
- Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the non-linear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4, 12-35.
- Presser, S., & Schuman, H. (1980). The measurement of a middle position in attitude surveys. *The Public Opinion Quarterly*, 44, 70-85.
- Raudenbush, S. W., & Bryk, A. G. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed). Thousand Oaks, CA: Sage Publications.
- Rugg, D., & Cantril, H. (1944). The wording of questions. In H. Cantril (Ed.), *Gauging Public Opinion*. Princeton, NJ: Princeton University Press.
- Samonte, K. (2011). *A longitudinal measurement invariance study of the sense of identity scale*. Unpublished manuscript.
- SAS Institute Inc., SAS® 9.2 Enhanced Logging Facilities [Computer software], Cary, NC: SAS Institute Inc., 2008.

- Saxon, D. P., Levine-Brown, P., Boylan, H. R. (2008). Affective assessment for developmental students, part 1. *Research in Developmental Education*, 22, 1-4.
- Si, S. X., & Cullen, J. B. (1998). Response categories and potential cultural bias: Effects of an explicit middle point in cross-cultural surveys. *International Journal of Organizational Analysis*, 6, 218-230.
- Spellings, M. (2006). *A test of leadership: Charting the future of U.S. higher education*. A report of the commission appointed by the Secretary of Education. Washington DC: U.S. Department of Education.
- Sudman, S., & Bradburn, N. M. (1974). *Response Effects in Surveys: A Review and Synthesis*. Chicago, Illinois: Aldine Publishing Company.
- Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update*, 14, 8-9.
- Thelk, A., Sundre, D. L., Horst, J. S., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale (SOS) to make valid inferences about student performance. *The Journal of General Education*, 58, 129-151.
- Tourangeau, R., Couper, M. P., & Conrad F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, 71, 91-112.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. New York, NY: Cambridge University Press.
- Velez, P., & Ashworth, S. D. (2007). The impact of item readability on the endorsement of the midpoint response in surveys. *Survey Research Methods*, 1, 69-74.
- Wang, Y., Ye, F., Jackson, G., Rodgers, R., & Jones, S. (2005). Development of student service-learning course survey (SSLCS) to measure service-learning course

outcomes. *IR Applications: Using Advanced Tools, Techniques, and Methodologies*, 3, 1-16.

Weng, L. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64, 956-972.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18, 163–83.

Zilberberg, A., Anderson, R. D., Finney, S. J., Marsh, K. R., & Swerdzewski, P. (2011, April). *Students' attitudes towards large-scale testing for institutional accountability mandates: Developing a measure*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Table 1

Fall 2012 Assessment Day Testing Session Configurations

Testing Configuration	Fall 2012 Assessment Administration Order					
	1	2	3	4	5	6
1			SDA-7		SOS	
2			SDA-7		SOS	
3						
4				SDA-7	SOS	
5						
6				SDA-7	SOS	
7			SDA-7		SOS	
8						
9						

Table 2

Study 1 SDA-7 SoI and M-GUDS Forms

SDA-7 Form	Item Numbers		Likert Response Option Scale							
	SoI	M-GUDS	1	2	3	4	5	6	7	
A	31-36	37-50	Strongly Disagree							Strongly Agree
B	31-36	37-50	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree	
C	31-36	37-50	Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree	
D	31-36	37-50	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree	
E	58-63	64-77	Strongly Disagree							Strongly Agree
F	58-63	64-77	Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree	
G	58-63	64-77	Strongly Disagree	Disagree	Slightly Disagree	Undecided	Slightly Agree	Agree	Strongly Agree	
H	58-63	64-77	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Agree	Strongly Agree	

Table 3

Study 1 Variable Intercorrelations and Descriptive Statistics

Variable	Neutral	Undecided	Neither Agree nor Disagree	Item Set Location	Effort	SAT Verbal	MRR
Neutral	--						
Undecided	--	--					
Neither Agree nor Disagree	--	--	--				
Item Set Location	-0.02	0.01	0.01	1.00			
Effort	0.00	-0.01	0.00	0.01	1.00		
SAT Verbal	0.00	-0.01	0.02	0.02	0.12**	1.00	
MRR	0.04	-0.02	-0.01	0.02	-0.18**	-0.05*	1.00
Mean	0.25	0.25	0.25	0.47	18.98	569.98	0.13
Standard Deviation	0.44	0.43	0.43	0.50	3.71	65.89	0.13
Skewness	1.13	1.15	1.18	0.12	-0.56	0.11	1.90
Kurtosis	-0.72	-0.68	-0.61	-1.99	0.54	0.07	6.81

Note. Neutral, Undecided, and Neither Agree nor Disagree are dummy-coded variables indicating midpoint label condition with the reference group being the unlabeled condition in which no label was provided for the midpoint. Intercorrelations amongst these midpoint label conditions were excluded within the table as they are not meaningful due to the experimental manipulation of these characteristics. Item set location is a dummy-coded variable indicating item set location condition with 0 and 1 corresponding to the earlier and later conditions, respectively. Although centered forms of Effort and SAT Verbal scores were used in the models, the uncentered forms of Effort and SAT Verbal were used to obtain the descriptive statistics shown here. Average MR is a respondent's average midpoint response across the 20 items. *denotes $p < .05$, **denotes $p < .01$.

Table 5

Study 1 Model 2 Results

Fixed Effect	Unstandardized Coefficient	SE	<i>p</i>
Intercept	-2.28	0.08	<0.001
Neutral	0.17	0.11	0.12
Undecided	0.09	0.11	0.44
Neither Agree nor Disagree	0.13	0.11	0.23
Item Set Location	0.23	0.11	0.04
Neutral*Item Set Location	-0.20	0.16	0.20
Undecided*Item Set Location	-0.21	0.16	0.18
Neither Agree nor Disagree*Item Set Location	-0.28	0.16	0.08

Table 6

Study 1 Model 3 Results

Fixed Effect	Unstandardized Coefficient	SE	<i>p</i>
Intercept	-2.29	0.08	<0.001
Neutral	0.17	0.11	0.11
Undecided	0.09	0.11	0.39
Neither Agree nor Disagree	0.15	0.11	0.17
Item Set Location	0.25	0.11	0.02
Effort	-0.05	0.02	0.01
Neutral*Item Set Location	-0.22	0.15	0.15
Undecided*Item Set Location	-0.24	0.16	0.12
Neither Agree nor Disagree*Item Set Location	-0.33	0.16	0.04
Item Set Location*Effort	0.01	0.03	0.73
Neutral*Effort	-0.01	0.03	0.63
Undecided*Effort	-0.01	0.03	0.83
Neither Agree nor Disagree*Effort	0.02	0.03	0.55
Item Set Location*Neutral*Effort	-0.01	0.04	0.85
Item Set Location*Undecided*Effort	0.01	0.04	0.86
Item Set Location*Neither Agree nor Disagree*Effort	-0.07	0.04	0.11

Table 7

Study 1 Model 4 Results

Fixed Effect	Unstandardized Coefficient	SE	<i>p</i>
Intercept	-2.290	0.078	<0.001
Neutral	0.177	0.107	0.099
Undecided	0.093	0.109	0.393
Neither Agree nor Disagree	0.141	0.110	0.200
Item Set Location	0.224	0.112	0.046
SAT Verbal	-0.002	0.001	0.053
Neutral*Item Set Location	-0.195	0.157	0.214
Undecided*Item Set Location	-0.208	0.158	0.187
Neither Agree nor Disagree*Item Set Location	-0.265	0.159	0.095
Item Set Location*SAT Verbal	-0.002	0.002	0.330
Neutral*SAT Verbal	0.004	0.002	0.013
Undecided*SAT Verbal	0.004	0.002	0.010
Neither Agree nor Disagree*SAT Verbal	0.001	0.002	0.565
Item Set Location*Neutral*SAT Verbal	0.000	0.002	0.896
Item Set Location*Undecided*SAT Verbal	-0.001	0.002	0.613
Item Set Location*Neither Agree nor Disagree*SAT Verbal	0.001	0.002	0.779

Table 8

Study 1 Exploratory Model 5 Results

Fixed Effect	Unstandardized Coefficient	SE	<i>p</i>
Intercept	-2.305	0.077	<0.0001
Neutral	0.184	0.106	0.082
Undecided	0.102	0.108	0.344
Neither Agree nor Disagree	0.159	0.109	0.145
Item Set Location	0.259	0.110	0.019
Effort	-0.046	0.021	0.028
SAT Verbal	-0.002	0.001	0.188
Neutral*Item Set Location	-0.221	0.154	0.152
Undecided*Item Set Location	-0.217	0.156	0.165
Neither Agree nor Disagree*Item Set Location	-0.315	0.157	0.045
Neutral*Effort	-0.018	0.029	0.539
Undecided*Effort	-0.021	0.029	0.481
Neither Agree nor Disagree*Effort	0.015	0.031	0.623
Neutral*SAT Verbal	0.004	0.002	0.022
Undecided*SAT Verbal	0.004	0.002	0.016
Neither Agree nor Disagree*SAT Verbal	0.000	0.002	0.766
Item Set Location*Effort	0.010	0.029	0.730
Item Set Location*SAT Verbal	-0.002	0.002	0.235
Effort*SAT Verbal	0.000	0.000	0.179
Neutral*Item Set Location*Effort	-0.013	0.041	0.753
Undecided*Item Set Location*Effort	0.016	0.041	0.707
Neither Agree nor Disagree*Item Set Location*Effort	-0.069	0.042	0.102
Neutral*Item Set Location*SAT Verbal	0.000	0.002	0.946
Undecided*Item Set Location*SAT Verbal	-0.001	0.002	0.705
Neither Agree nor Disagree*Item Set Location*SAT Verbal	0.001	0.002	0.604
Neutral*Effort*SAT Verbal	-0.002	0.000	0.001
Undecided*Effort*SAT Verbal	0.000	0.000	0.712
Neither Agree nor Disagree*Effort*SAT Verbal	0.000	0.000	0.657
Item Set Location*Effort*SAT Verbal	0.000	0.000	0.290
Neutral*Item Set Location*Effort*SAT Verbal	0.001	0.001	0.021
Undecided*Item Set Location*Effort*SAT Verbal	0.000	0.001	0.823
Neither Agree Nor Disagree*Item Set Location*Effort*SAT Verbal	0.000	0.001	0.745

Table 9

Study 1 Exploratory Model 6 Results

Fixed Effect	Unstandardized Coefficient	SE	<i>p</i>
Intercept	-2.243	0.067	<0.0001
Neutral	0.120	0.099	0.224
Undecided	-0.015	0.077	0.847
Neither Agree nor Disagree	0.090	0.100	0.367
Item Set Location	0.142	0.078	0.068
Effort	-0.049	0.012	<0.0001
SAT Verbal	-0.001	0.001	0.158
Neutral*Item Set Location	-0.103	0.133	0.438
Neither Agree nor Disagree*Item Set Location	-0.176	0.134	0.189
Neutral*Effort	-0.014	0.023	0.533
Neutral*SAT Verbal	0.003	0.001	0.016
Undecided*SAT Verbal	0.003	0.001	0.003
Item Set Location*Effort	-0.006	0.017	0.726
Item Set Location*SAT Verbal	-0.002	0.001	0.056
Effort*SAT Verbal	0.000	0.000	0.125
Neutral*Item Set Location*Effort	0.003	0.034	0.938
Neutral*Item Set Location*SAT Verbal	0.000	0.002	0.865
Neutral*Effort*SAT Verbal	-0.001	0.000	0.000
Item Set Location*Effort*SAT Verbal	0.000	0.000	0.130
Neutral*Item Set Location*Effort*SAT Verbal	0.001	0.001	0.010

Table 10

Study 1 Fit Statistics, τ_{00} for Models 2-6, and Likelihood Ratio Test Results Comparing Models 2, 3, 4, and 6 to Model 5

Model	τ_{00}	Deviance (-2LL)	AIC	BIC	Number of Parameters	df	χ^2	p
Model 5	0.77	27109	27175	27357	33			
Model 6	0.78	27119	27161	27277	21	12	10	0.62
Model 4	0.83	27189	27223	27316	17	16	80	<0.001
Model 3	0.79	27145	27179	27280	17	16	36	0.003
Model 2	0.85	27213	27231	27280	9	24	104	<0.001

Table 11

Study 2 Think-aloud Coding Scheme

Organizing Theme	Basic Theme	Coding	Example response(s)
	Ambivalence	In the middle/ambivalence	“I would choose four because I don’t think it really has a beneficial or negative impact it just is what it is.”; “I didn’t necessarily agree or disagree...”; “Four - I neither agree nor disagree. Because some connections may be positive, some connections may be negative. I guess it really just depends on what those differences are...”
		Noncommittal	“I neither agree nor disagree because whether I know how they’re different from me doesn’t really help me get along with them...”
	Non-commitment	Slightly agree	“Four – I’m neutral. I do like to meet people from different ethnic groups, like I had a best friend that’s from...I do like meeting from different ethnic groups. It does kind-of expand my horizons to know more things.”
		Slightly disagree	“I feel like maybe neutral on that one because I feel like people always have hidden things that they don’t necessarily open up about or keep to themselves...”
Indecision	Uncertainty	Don't know/Not applicable	“I would have to pick number four because I don’t know how it relates to myself or how it doesn’t – it seems good if you were to look at people in general and how they react to each other because of different ethnicities...to look at it as a part of yourself is kind of abstract.”
		Undecided	“Well, I’m kind of in-between, undecided on it.”
	Indifference	Indifference	“Kind of, but also kind of not. We’ll say four because I’m kind of indifferent about that one.”
		No opinion	“I would probably pick no opinion. Cause you know, you...it just depends on how you feel about that subject too.”
	Item statement clarity	Confusion	“I didn’t necessarily agree or disagree, and at the same time I didn’t completely understand what it’s asking for.”; “Not entirely sure what this one is asking. I think maybe I know where I fit in the world, but I also don’t...so I’ll say four – neither agree nor disagree, I suppose.”

Note. Shading indicates basic themes encompassed within *Indecision* organizing theme.

Table 12

Study 2 Think-aloud Midpoint Endorsement Summary

Item Number					Frequency Endorsing	
SoI	M-GUDS	Think-aloud Form A	Item Statement	Midpoint Label	Midpoint	Participant
1		1	I have a definite sense of purpose in life.	unlabeled	0	
	12 (SELFUND)	2	Knowing about the experiences of people of different races increases my self-understanding.	unlabeled	4	3,5,6,7
	2 (SIMDIF)	3	In getting to know someone, I like knowing both how he/she differs from me and is similar to me.	unlabeled	1	8
2		4	I know what I want out of life.	<i>Neutral</i>	0	
	4 (SELFUND)	5	Knowing someone from a different ethnic group broadens my understanding of myself.	<i>Neutral</i>	4	1,4,7,8
	11 (SIMDIF)	6	I can best understand someone after I get to know how he/she is both similar and different from me.	<i>Neutral</i>	1	6
3		7	I have a clear set of personal values or moral standards.	<i>Undecided</i>	0	
	5 (SELFUND)	8	Knowing about the different experiences of other people helps me understand my own problems better.	<i>Undecided</i>	0	
	14 (SIMDIF)	9	In getting to know someone, I try to find out how I am like that person as much as how that person is like me.	<i>Undecided</i>	1	8
4*		10	I don't know where I fit in the world.	<i>Neither Agree nor Disagree</i>	2	5, 6
5		11	I have specific personal goals for the future.	<i>Neither Agree nor Disagree</i>	0	
	7 (SIMDIF)	12	Knowing how a person differs from me greatly enhances our friendship.	<i>Neither Agree nor Disagree</i>	2	4, 7

Note. * denotes item reverse-scoring.

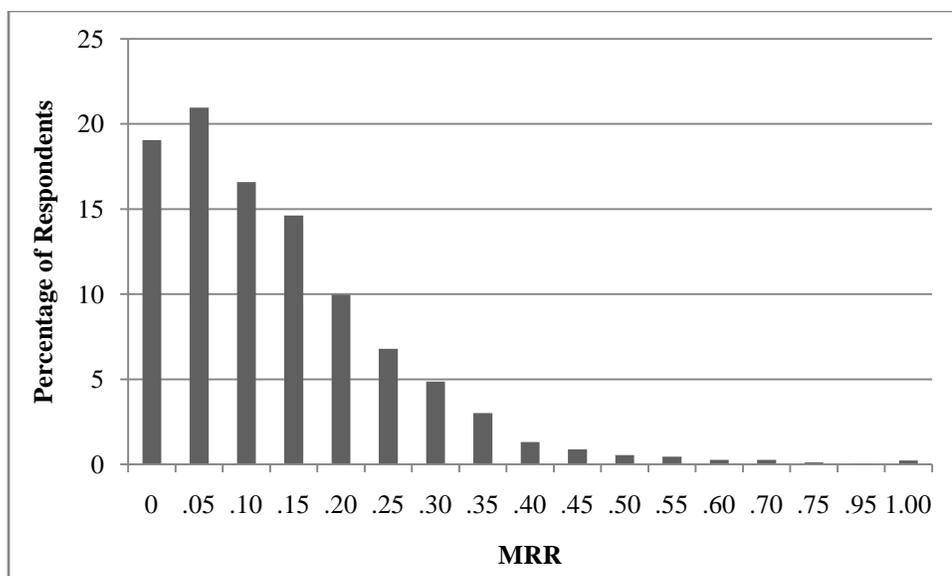


Figure 1. Histogram of average midpoint response endorsement across items for the 1,826 respondents

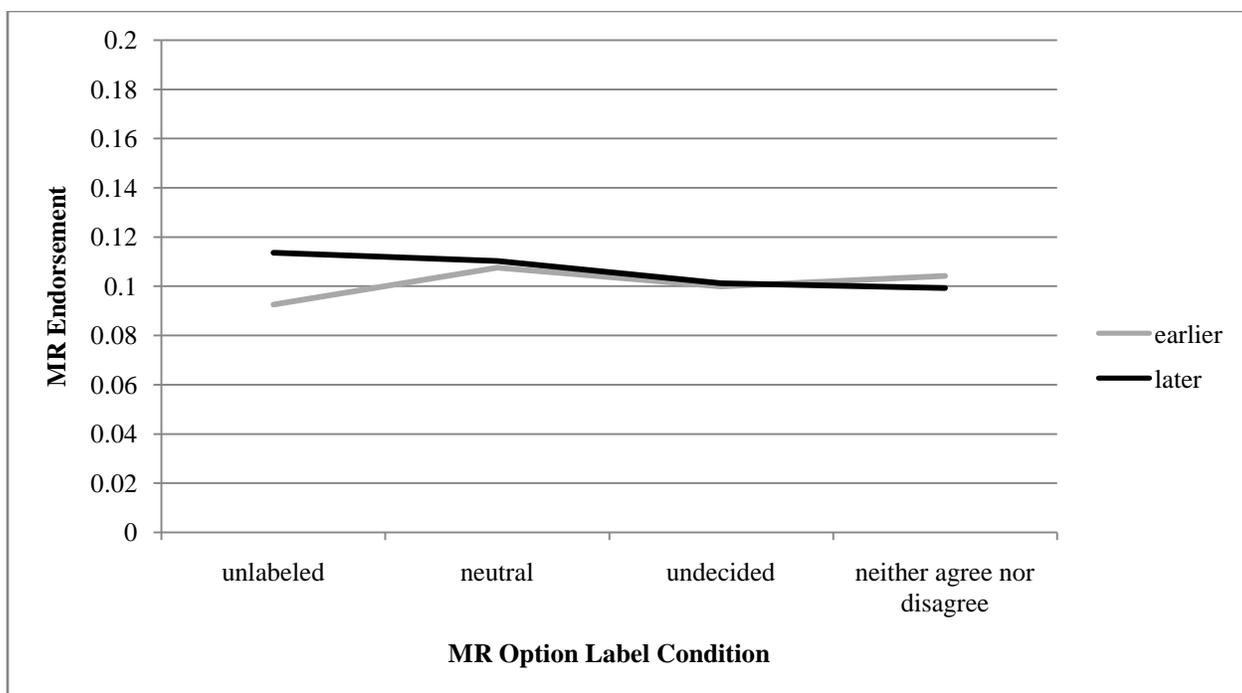


Figure 2. Predicted probability of midpoint response (MR) endorsement across model 2 predictor combinations

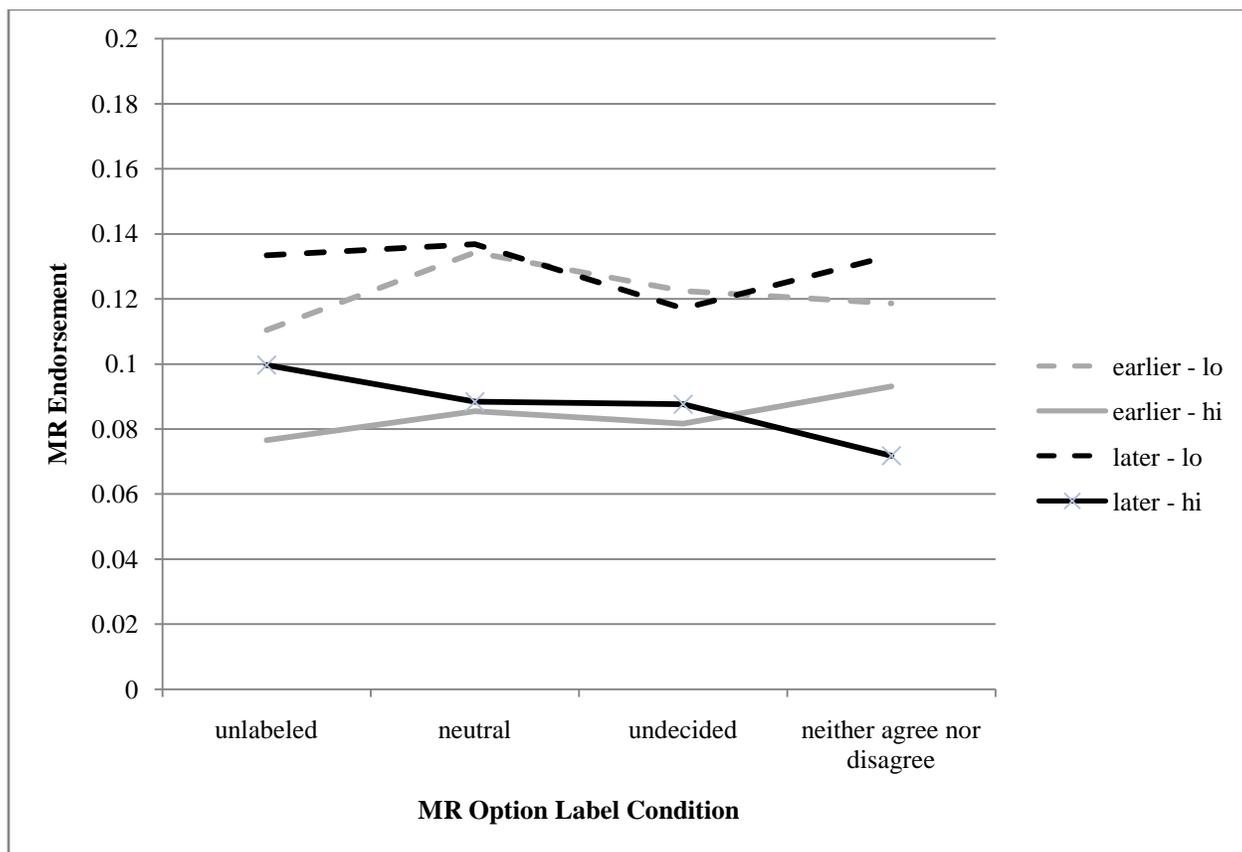


Figure 3. Predicted probability of midpoint response (MR) endorsement across model 3 predictor combinations

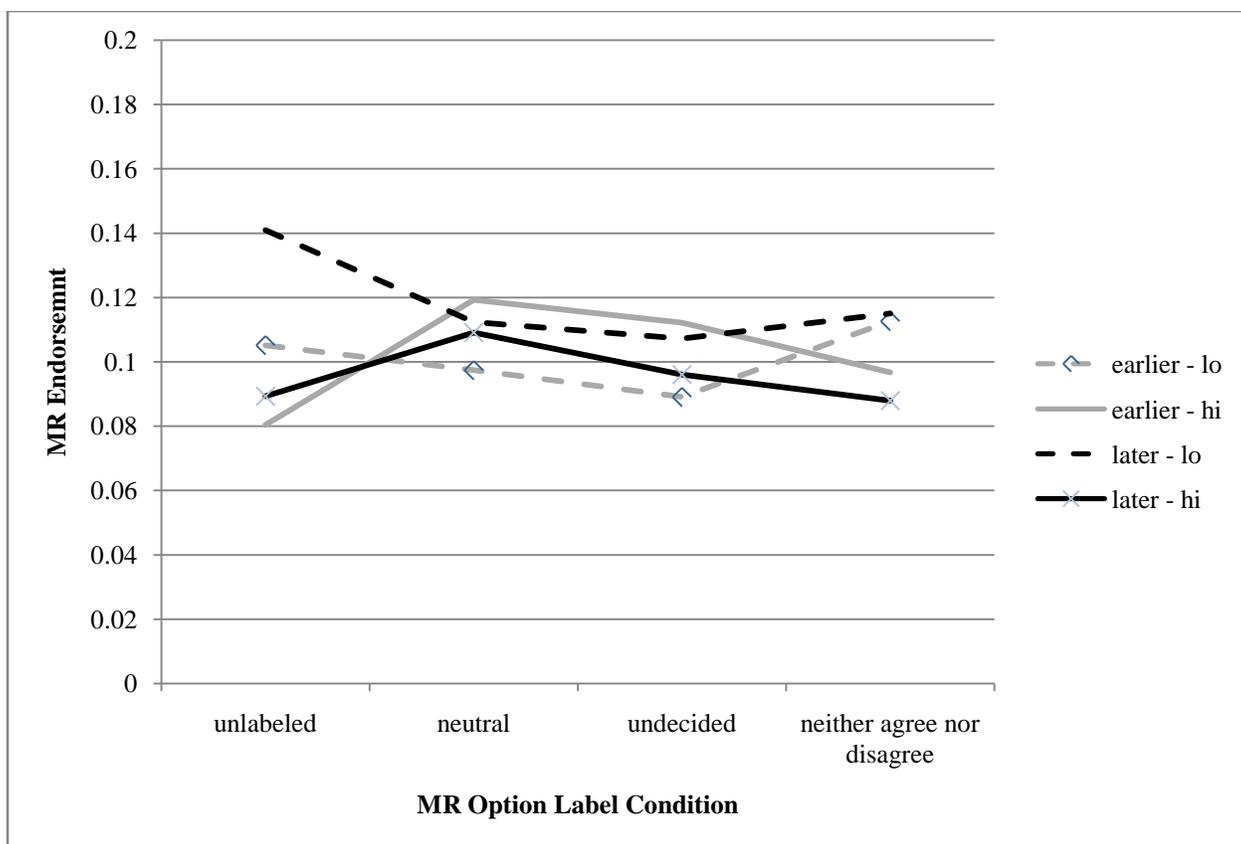


Figure 4. Predicted probability of midpoint response (MR) endorsement across model 4 predictor combinations

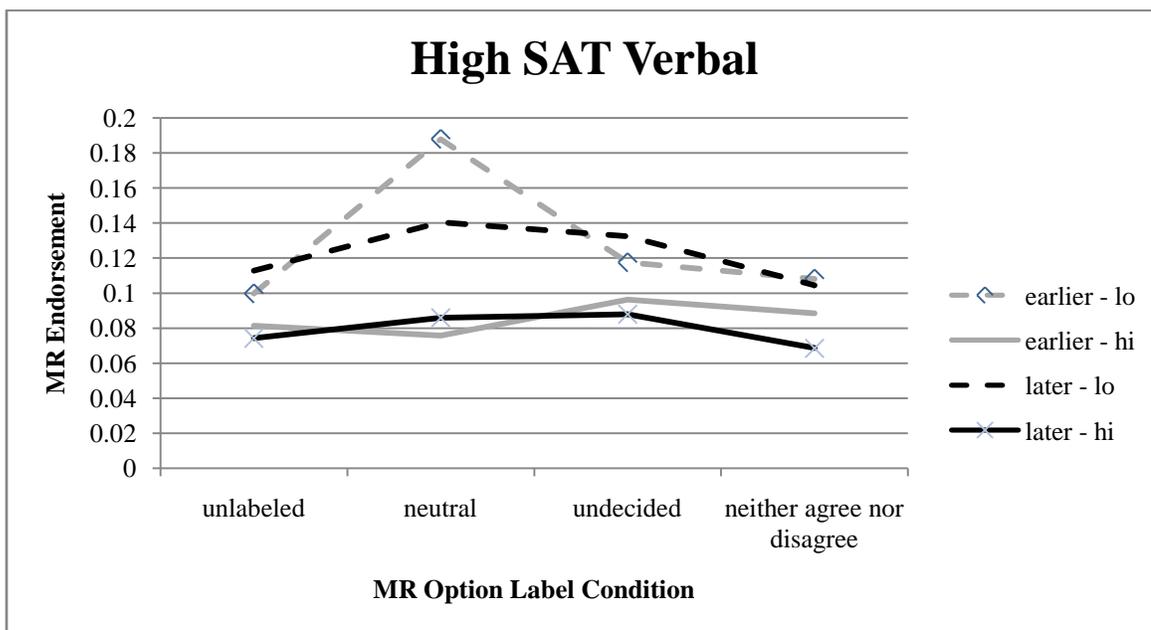


Figure 5. Predicted probability of midpoint response (MR) endorsement for high SAT verbal respondents across model 6 predictor combinations

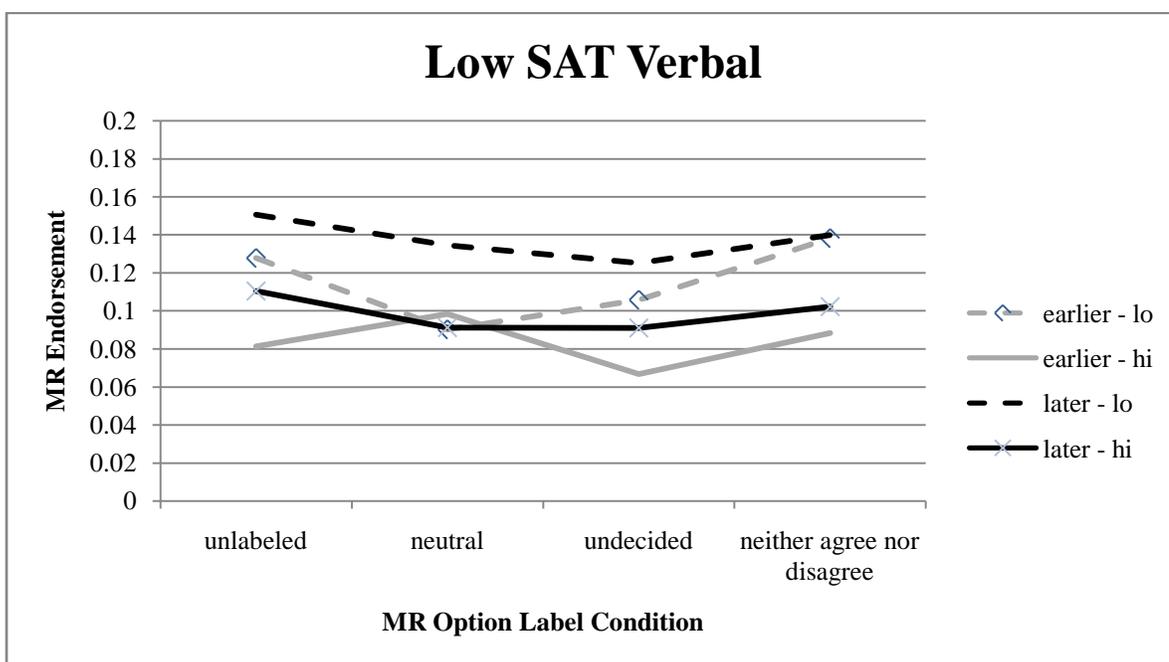


Figure 6. Predicted probability of midpoint response (MR) endorsement for low SAT verbal respondents across model 6 predictor combinations