A PSYCHOMETRIC INVESTIGATION OF THE CONCEPTUAL STRUCTURE
OF THE REVISED NATIONAL SURVEY OF STUDENT ENGAGEMENT (NSSE 2.0)

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY
OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

EDUCATIONAL PSYCHOLOGY

August 2021

By

Jessica L. W. Miranda

Dissertation Committee:

Seongah Im, Chairperson
Min Liu
George Harrison
Ron Heck
Stacey Roberts

Keywords: NSSE, psychometrics, IRT, CFA
Dedication

This work is dedicated to my son, Seth, who gave me the strength to persevere and whose bright mind inspires me every day; to my husband, Emmanuel, who has always encouraged me in pursuing my goals; and to my mom, Jenny, who introduced me to academia as a child and who has been a steadfast mentor.
Acknowledgements

The guidance, mentoring, and patience with my innumerous questions that my
dissertation chair, Dr. Seongah Im, has provided me have been invaluable in reaching this
accomplishment. Dr. Im goes above and beyond to help her students understand the complexities
of psychometrics and her knowledge base is enviable. Dr. Im, you are an exemplary advisor and
I am blessed to have had the honor of working with you. In addition, Dr. Min Liu, Dr. George
Harrison, and Dr. Ron Heck have all taught me so much in regards to statistics, measurement,
and evaluation; I am forever grateful for your support during my doctoral studies and dissertation
process. Lastly, Dr. Stacey Roberts has never failed to be an incredibly supportive mentor and a
smiling face to count on. Thank you all so much for helping me to join you in the ranks of
doctoral level faculty.
Abstract

Student engagement plays an important role in student learning and success in elementary through postsecondary education. There has been growing interest in student engagement and its influence on student persistence, retention, and overall success in education over the past several decades. The importance of engagement has led researchers and university administrators to seek out measures of engagement that will provide information for improving undergraduate student success and that provide a means of assessing educational quality. The National Survey of Student Engagement (NSSE) was developed in the hopes of providing valid and reliable information about institutional quality and student engagement, and is administered to hundreds of thousands of freshmen and seniors each year at over 500 institutions. In 2013, the NSSE underwent a major revision; however, the psychometric properties of the revised instrument (i.e., NSSE 2.0) have not been thoroughly investigated and validated for the instrument’s intended uses and score interpretations.

This study addressed the dearth of research investigating the psychometric properties and validity evidence of the revised NSSE 2.0, including the validity of the proposed conceptual structure for measuring student engagement. The findings of the current study partially support that the NSSE 2.0 has adequate psychometric quality to make confident decisions utilizing the NSSE engagement indicator scores at the institution level. The Academic Challenge and Learning with Peers themes were deemed adequate psychometrically for their intended use; however, the findings for the Experiences with Faculty theme were inconclusive and the Campus Environment theme was lacking in psychometric quality. At the item level, the majority of the engagement indicator items appeared to be working well based on item parameter estimates and information functions, though there was local dependency among items of the engagement
indicators. Both the CFA and IRT results indicated that there was a multidimensional or higher
order nature to the NSSE engagement indicators. However, due to mixed results across
engagement indicators and themes and issues with goodness of fit of numerous models, the
NSSE’s suggested model structure of the construct of student engagement with ten engagement
indicators organized into four themes was not fully confirmed.
Table of Contents

Acknowledgements ........................................................................................................................ iv
Abstract ........................................................................................................................................... v
List of Tables ................................................................................................................................. xi
List of Figures ............................................................................................................................... xii
Chapter 1: Introduction ................................................................................................................... 1
  Significance of Study .................................................................................................................. 2
  Purpose of Study ......................................................................................................................... 3
Chapter 2: Literature Review .......................................................................................................... 4
  Theories of Student Engagement ................................................................................................. 4
    Theoretical approach of study. .................................................................................................. 15
  Defining Student Engagement for the College-Age Population ............................................. 16
  Measuring Student Engagement for the College-Age Population ........................................... 20
    Person-Oriented: End of continuum. ..................................................................................... 20
    Person-in-Context: Middle of continuum. ............................................................................. 22
    Context-Oriented: End of continuum. ................................................................................... 24
  Validation Studies of Student Engagement Measures .............................................................. 25
    The National Survey of Student Engagement. ...................................................................... 26
    The Student Experience in the Research University Survey. ................................................. 34
    The National Student Survey................................................................................................ 36
  Item Response Theory: Unidimensional and Multidimensional Examination ....................... 39
  Confirmatory Factor Analysis: First-Order, Bifactor, and Higher-Order Factor Structures..... 42
  Summary ................................................................................................................................... 47
Chapter 3: Methodology ............................................................................................................... 49
  Data Source ................................................................................................................................ 49
  Data Cleaning ............................................................................................................................ 53
  Data Analysis ............................................................................................................................ 53
    Model Evaluation .................................................................................................................. 57
    Exploratory Factor Analysis ................................................................................................. 59
    Item Response Theory .......................................................................................................... 59
    Confirmatory Factor Analysis ............................................................................................ 60
      Testing for Configural Invariance ..................................................................................... 60
First Order Confirmatory Factor Analysis ......................................................................... 60
Second-Order Confirmatory Factor Analysis .................................................................... 64
Bifactor Confirmatory Factor Analysis ............................................................................. 68
Model Comparison ............................................................................................................. 72
Chapter 4: Results ......................................................................................................................... 74
Exploratory Factor Analysis ...................................................................................................... 74
All 47 Items ........................................................................................................................... 74
Academic Challenge Theme .................................................................................................. 77
Learning with Peers Theme ................................................................................................... 78
Experiences with Faculty Theme .......................................................................................... 78
Campus Environment Theme ................................................................................................ 79
Item Response Theory Analysis ................................................................................................. 80
Academic Challenge Theme ................................................................................................ 80
Unidimensional IRT of Higher Order Learning Engagement Indicator ......................... 80
Unidimensional IRT of Reflective and Integrative Learning Engagement Indicator .......... 83
Unidimensional IRT of Quantitative Reasoning Engagement Indicator ............................ 86
Unidimensional IRT of Learning Strategies Engagement Indicator .................................. 88
MIRT of Academic Challenge Theme ............................................................................... 90
Learning with Peers Theme ................................................................................................ 92
Unidimensional IRT of Collaborative Learning Engagement Indicator ............................ 92
Unidimensional IRT of Discussions with Diverse Others Engagement Indicator ............... 94
MIRT of Learning with Peers Theme ................................................................................ 97
Experiences with Faculty Theme .......................................................................................... 98
Unidimensional IRT of Student-Faculty Interactions Engagement Indicator .................. 98
Unidimensional IRT of Effective Teaching Practices Engagement Indicator .................... 101
MIRT of Experiences with Faculty Theme ........................................................................ 104
Campus Environment Theme .............................................................................................. 105
Unidimensional IRT of Quality of Interactions Engagement Indicator ............................ 105
Unidimensional IRT of Supportive Environment Engagement Indicator ......................... 108
MIRT of Campus Environment Theme ............................................................................ 111
Confirmatory Factor Analysis ............................................................................................... 112
Academic Challenge Theme ............................................................................................... 112
List of Tables

Table 1 NSSE 2.0 Engagement Indicator and Theme Structure ................................................. 29
Table 2 NSSE 2.0 Engagement Indicator Survey Items ............................................................. 50
Table 3 All 47 NSSE Items Factor Loadings of EFA ............................................................... 75
Table 4 Academic Challenge Theme Factor Loadings of EFA ............................................... 77
Table 5 Learning with Peers Theme Factor Loadings of EFA .............................................. 78
Table 6 Experiences with Faculty Theme Factor Loadings of EFA ....................................... 79
Table 7 Campus Environment Theme Factor Loadings of EFA .......................................... 80
Table 8 G factor and Unique Factor Contributions in Academic Challenge Bifactor Model... 119
Table 9 Learning with Peers Bifactor Model Comparison of G factor and Unique Factor Contributions .......................................................................................................................... 123
Table 10 Experiences with Faculty Bifactor Model Comparison of G factor and Unique Factor Contributions .......................................................................................................................... 128
List of Figures

Figure 1 Kuh, Kinzie, Buckley, Bridges, and Hayek’s (2007b) Framework, “What Matters to Student Success” .......................................................................................................................................................................................................................................................... 9

Figure 2 Kahu’s (2013) Conceptual Framework of Engagement, Antecedents and Consequences .......................................................................................................................................................................................................................................................... 12

Figure 3 Kahu and Nelson’s (2017) Refined Conceptual Framework of Student Engagement Incorporating the Educational Interface .......................................................................................................................................................................................................................................................... 14

Figure 4 Examples of Simple Unidimensional and Multidimensional Models .......................................................................................................................................................................................................................................................... 41

Figure 5 Illustration of Second-Order and Bifactor Model Structures .......................................................................................................................................................................................................................................................... 45

Figure 6 Data Analysis Flowchart .......................................................................................................................................................................................................................................................................................................................... 55

Figure 7 Academic Challenge CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR) .......................................................................................................................................................................................................................................................................................................................................................................................... 61

Figure 8 Learning with Peers CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD) .......................................................................................................................................................................................................................................................................................................................................................................................... 62

Figure 9 Experiences with Faculty CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET) .......................................................................................................................................................................................................................................................................................................................................................................................... 63

Figure 10 Campus Environment CFA Model: Quality of Interactions (QI) and Supportive Environment (SE) .......................................................................................................................................................................................................................................................................................................................................................................................... 64

Figure 11 Academic Challenge (AC) 2nd Order CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR) ........................................ 65

Figure 12 Learning with Peers (LP) 2nd Order CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD) .......................................................................................................................................................................................................................................................................................................................................................................................... 66
Figure 13 Experiences with Faculty (EF) 2nd Order CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET) ................................................................. 67

Figure 14 Campus Environment (CE) 2nd Order CFA Model: Quality of Interactions (QI) and Supportive Environment (SE) .......................................................................................... 68

Figure 15 Academic Challenge (AC) Bifactor CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR) .... 69

Figure 16 Learning with Peers (LP) Bifactor CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD) .................................................................................. 70

Figure 17 Experiences with Faculty (EF) Bifactor CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET) .................................................................................. 71

Figure 18 Campus Environment (CE) Bifactor CFA Model: Quality of Interactions (QI) and Supportive Environment (SE) .......................................................................................... 72

Figure 19 Category Response Curves and Information Functions for Higher Order Learning EI ................................................................................................................................. 82

Figure 20 Total Information Curve for Higher Order Learning EI ....................................................... 83

Figure 21 Category Response Curves and Information Functions for Reflective and Integrative Learning EI ................................................................................................................................. 85

Figure 22 Total Information Curve for Reflective and Integrative Learning EI ............................................. 85

Figure 23 Category Response Curves and Information Functions for Quantitative Reasoning EI ................................................................................................................................. 87

Figure 24 Total Information Curve for Quantitative Reasoning EI ....................................................... 88

Figure 25 Category Response Curves and Information Functions for Learning Strategies EI... 90

Figure 26 Total Information Curve for Learning Strategies EI ................................................................. 90
Figure 27 Category Response Curves and Information Functions for Collaborative Learning EI ................................................................. 94

Figure 28 Total Information Curve for Collaborative Learning EI ................................................................. 94

Figure 29 Category Response Curves and Information Functions for Discussion with Diverse Others EI ........................................................................................................................................ 96

Figure 30 Total Information Curve for Discussions with Diverse Others EI .................................................. 97

Figure 31 Category Response Curves and Information Functions for Student-Faculty Interactions EI ........................................................................................................................................ 100

Figure 32 Total Information Curve for Student-Faculty Interactions EI ................................................................. 101

Figure 33 Category Response Curves and Information Functions for Effective Teaching Practices EI ........................................................................................................................................ 103

Figure 34 Total Information Curve for Effective Teaching Practices EI ................................................................. 103

Figure 35 Category Response Curves and Information Functions for Quality of Interactions EI ........................................................................................................................................ 107

Figure 36 Total Information Curve for Quality of Interactions EI ................................................................. 108

Figure 37 Category Response Curves and Information Functions for Supportive Environment EI ........................................................................................................................................ 110

Figure 38 Total Information Curve for Supportive Environment EI ................................................................. 110

Figure 39 Academic Challenge 1st Order CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Quantitative Reasoning (QR), Learning Strategies (LS) .... 114

Figure 40 Academic Challenge (AC) 2nd Order CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR) .. 116
Figure 41 Academic Challenge Bifactor CFA Model: Higher-Order Learning (HO), Reflective & Integrative Learning (RI), Quantitative Reasoning (QR), and Learning Strategies (LS) ..... 118

Figure 42 Learning with Peers CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD).......................................................................................................................... 121

Figure 43 Learning with Peers (LP) Bifactor CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD).......................................................................................................................... 123

Figure 44 Experiences with Faculty CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET) .......................................................................................................................... 126

Figure 45 Experiences with Faculty (EF) Bifactor CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET) .......................................................................................................................... 128

Figure 46 Campus Environment CFA Model: Quality of Interactions (QI) and Supportive Environment (SE).......................................................................................................................... 131
Chapter 1: Introduction

Student engagement plays an important role in student learning and success in elementary through postsecondary education. The level at which students engage in their educational experiences ultimately influences their likelihood for success and persistence (Kuh et al., 2006). There has been growing interest in student engagement and its influence on student persistence, retention, and overall success in education over the past several decades (Trowler, 2010). The importance of engagement has led researchers and university administrators to seek out measures of engagement that will provide information for improving undergraduate student success and that provide a means of assessing educational quality (Campbell & Cabrera, 2011; Miller & Malandra, 2006). Despite the increased research on student engagement, there is considerable variation in the definition and measurement of student engagement across the field (Fredricks & McColskey, 2012; Henrie et al., 2015; Kahu, 2013; Sinatra et al., 2015; Zepke, 2015).

The National Survey of Student Engagement (NSSE) was developed in the hopes of providing valid and reliable information about institutional quality and student engagement. The NSSE, which is also called “The College Student Report,” is one of the most widely used measures of student engagement in North America with 531 institutions and 294,507 freshman and senior undergraduate students participating in 2019 (The Center for Postsecondary Research, 2019). The NSSE defines the construct of student engagement as, “the amount of time and effort students put into their studies and other educationally purposeful activities” (The Center for Postsecondary Research, 2017, para. 1). In 2013, the NSSE underwent a major revision; however, the psychometric properties of the instrument have not been thoroughly investigated and validated for the instrument’s intended uses and score interpretations.
Significance of Study

This study addressed the dearth of research exploring whether the uses and interpretations of the revised NSSE (i.e., NSSE 2.0) scores are psychometrically sound. Only one published study (Carle et al., 2009) was found that utilized item response theory to analyze the NSSE. This study utilized NSSE data prior to 2013 when the instrument underwent a significant revision. Research has shown that changes to the content of a measurement instrument can have significant influences on the psychometric properties of the instrument, and therefore when an instrument undergoes major revisions, its psychometric properties must be reexamined to ensure that the uses and interpretation of scores are psychometrically appropriate (Furr, 2011).

In addition, the dearth of research investigating the validity of the proposed model structure for measuring student engagement of NSSE 2.0 needs to be addressed. As NSSE 2.0 is relatively new, there were relatively few published validation studies on the new 2013 version. Fosnacht and Gonyea (2018) examined the generalizability of the NSSE 2.0 engagement indicators with national data and found that the engagement indicators can reliably be generalized from small samples to a larger population of students; however, they also found that only some of the engagement indicators can dependably be generalized to higher-order constructs using small samples. Two validation studies of NSSE 2.0 (Le, 2019; Winkler, 2020) on single institutions’ data raised concerns about the reliability and validity of the NSSE’s ten engagement indicators and four themes on their data sets. Recent studies (Campbell & Cabrera, 2011; Gordon et al., 2008; LaNasa et al., 2009; Tendhar et al., 2013) of institutional level NSSE data for single, public, research institutions prior to its 2013 revision have found that the NSSE benchmarks (latent construct) model was a poor representation for their data. Based upon their findings, LaNasa et al. (2009) found that "the strongest implication is a cautionary one: It is
incumbent on institutions to fully explore their own data, especially when using the data for comparative purposes. To fully understand NSSE results, item level inspection is required" (p.330).

**Purpose of Study**

Through an in-depth psychometric analysis of the University of Hawai‘i at Mānoa’s (UHM) 2015 NSSE data, this study addressed the research gap and contributed to the field of measurement by addressing the following research questions: (a) Based on unidimensional item response theory (IRT) and multidimensional item response theory (MIRT) criteria, are the NSSE items of the ten engagement indicators psychometrically sound?; (b) Based on confirmatory factor analysis (CFA) criteria, does the NSSE’s suggested factor structure of the construct of student engagement with ten engagement indicators organized into four themes fit the UHM 2015 NSSE data, and does a bifactor or second-order factor structure display a better fit?; and (c) What do the integrated results of (a) and (b) reveal about the psychometric adequacy of the NSSE 2.0?
Chapter 2: Literature Review

Theories of Student Engagement

The research on student engagement has expanded significantly since the 1980s. However, despite the growing research into student engagement, there is no overarching theory of engagement (Azevedo, 2015; Boekaerts, 2016). Azevedo (2015) argued that “there is a general lack of consistency in the specification and articulation of the theoretical underpinnings of the construct” (p. 85). While there is plurality in the conceptual and theoretical underpinnings of student engagement across the field (Kahu, 2013; Zepke, 2015), much of the research on student engagement within higher education has its roots in the seminal works of Astin (1984) and Tinto (1975).

Astin’s (1984) theory of student involvement was a foundational theory to the study of student engagement. Based on a longitudinal study of college dropouts, Astin (1975) found that all significant factors for persistence in college were related to student involvement. Student involvement according to Astin (1999) was “the amount of physical and psychological energy that the student devotes to the academic experience” (p. 518). The most significant environmental factors for college persistence in Astin’s (1975) study were: (a) student's residence on or off-campus (on-campus living was positively related to persistence for all institution types and demographic groups), (b) extracurricular activities and membership in Greek clubs (i.e., sororities and fraternities), (c) having a part-time job on-campus (positive effect) versus having a full-time off campus job (negative effect), (d) institution type (two year colleges were a negative factor in comparison to four year colleges), and (e) "fit" between the student and the college environment (students were more likely to be involved, and persist, if they identified with the college environment, such as through religion or race/ethnicity). Across

4
the literature, “fit” is often referred to as sense of belonging, which is defined as a "students’
sense of connection with their college, degree of social support, and experience of both academic
challenge and support" (Glass & Westmont, 2014, p. 108). Higher education research in the
United States and United Kingdom has suggested that when students have developed a sense of
belonging to the educational environment, they will be more likely to engage (Glass &
Westmont, 2014; Hughes, 2010; McCune, 2009). Astin (1984) acknowledged that there were
internal aspects to involvement; however, he believed that the most critical and important aspects
of student involvement were behavioral. This focus on the behavioral perspective of student
involvement has carried through into the majority of research on student engagement over the
last three decades (Kahu, 2013; Zepke, 2015).

Another seminal work for student engagement was Tinto's (1975, 1987) theory of student
departure which was foundational to the focus on the emotional or affective aspect of student
engagement (Zepke, 2015). According to Tinto’s theory, students who socially and academically
integrate into the campus community increase their commitment to the institution and are more
likely to graduate. The degree of students’ academic and social integration is influenced by their
level of commitment to their educational goals and college. Their level of integration in turn
influences their degree of continued commitment and persistence in college. The process of
integration involves three stages: (a) separation from norms of prior life, (b) transition towards
the adoption of new norms, and (c) incorporation into norms of the higher education institution,
which relates to the extent students perceive they "fit" within the educational environment
(Tinto, 1975).

Many studies have found that undergraduates’ involvement with peers and faculty
influences their identity development and is associated with students’ engagement in higher
education (Glass & Westmont, 2014; McCune, 2009; Milem & Berger, 1997; Zepke & Leach, 2010). In a study of a conceptual model that brought together the behavioral aspect of Astin’s (1984) theory of involvement and the integration process of Tinto’s (1975) theory of student departure, Milem and Berger (1997) examined the influences of peer and faculty involvement on students’ integration. Milem and Berger found that students more involved with peers were more likely to have higher levels of academic integration, social integration, and institutional commitment; students more involved with faculty were more likely to have higher levels of academic integration; and that social integration, though not academic integration, was a significant positive predictor of institutional commitment and intent to reenroll. Through an extensive research synthesis on student engagement studies, Zepke and Leach (2010) also concluded that educational experiences that foster collaborative involvement with peers and faculty can lead to increases in student engagement.

According to Trowler (2010), “engagement is more than involvement or participation” as it involves cognition and emotions in addition to activity (i.e., behaviors) (p. 5). Within the engagement literature, there are several different perspectives or conceptualizations of engagement used by scholars which take behavioral, emotional, cognitive, and/or holistic approaches to student engagement (Fredricks & McColskey, 2012; Trowler, 2010). Fredricks et al. (2004) proposed that engagement is a multidimensional or “meta” construct and put forward the tripartite model of engagement (i.e., behavioral, emotional, and cognitive dimensions of engagement). In this light, student engagement is a meta-construct that includes, “behavioral (i.e., time on task), emotional (i.e., interest and value), and cognitive engagement (i.e., self-regulation and learning strategies)” (Fredricks & McColskey, 2012, p. 764). The tripartite model was focused on engagement in school-aged children (i.e., elementary and secondary); however, it
has since been utilized extensively within higher education research. Fredricks et al. (2004) called for the differentiation of the three types of engagement, as well as the specification of antecedents, which they specified as school-level factors, classroom context including teacher and peer support/relationships, classroom structure, task characteristics, and individual needs for relatedness, autonomy, and competence; as well as the specification of outcomes including achievement and dropping out. According to Reschly and Christenson (2012) the tripartite model proposed by Fredricks et al. “was the elevation of engagement to the level of a metaconstruct that brings together many previously separate lines of research and may subsume the construct of motivation as a form of engagement” (p. 11).

The research on engagement has “grown out of a variety of different theoretical traditions” (Fredricks et al., 2016, p. 1) and, because of this, the concepts within the three types of engagement overlap with many constructs that have been studied in separate lines of research. For example, Fredricks et al. (2004) argued that,

research on behavioral engagement is related to that on student conduct and on-task behavior (Karweit, 1989; Peterson, Swing, Stark, & Wass, 1984). Research on emotional engagement is related to that on student attitudes (Epstein & McPartland, 1976; Yamamoto, Thomas, & Karns, 1969) and student interest and values (Eccles et al., 1983). Research on cognitive engagement is related to that on motivational goals and self-regulated learning (Boekarts, Pintrich, & Zeidner, 2000; Zimmerman, 1990). (p. 60)

Accordingly, scholars have drawn upon a variety of theoretical frameworks when studying student engagement including theories related to motivation, such as self-determination theory, self-regulated learning theory, expectancy-value theory, flow theory, and goal theory; as well as school identification and connection theories to understand engagement within the process of
student attrition and/or completion (Boekaerts, 2016; Eccles, 2016; Fredricks et al., 2016).

Some researchers have drawn upon these diverse theories to expand upon Fredricks et al.’s tripartite model (Boekaerts, 2016). Reeve and Tseng (2011) drew upon Ryan and Deci’s (2000) self-determination theory when they proposed the addition of agentic engagement (i.e., students’ constructive contribution to instruction) to Fredricks et al.’s tripartite model. Within self-determination theory, students’ psychological needs of autonomy, competence, and relatedness are motivating assets within the learning process (Jang et al., 2016). While from the self-determination theory perspective, adding a separate fourth dimension for agentic engagement makes sense, scholars working from a different theoretical perspective may reject this addition based upon the argument that agency within the learning process should fall within the cognitive dimension of engagement (Boekaerts, 2016).

While there are some arguments for additional dimensions of engagement, there is general consensus across the field that engagement is a multidimensional construct (Boekaerts, 2016). Engagement is also often viewed as context dependent or malleable across educational and social situations (Boekaerts, 2016; Cleary & Zimmerman, 2012; Kahu & Nelson, 2017). Cleary and Zimmerman (2012) argued that “the extent to which students become cognitively or strategically immersed in learning activities will vary across academic settings and situational demands” (p. 240). Similarly, Boekaerts (2016) proposed that “student engagement…is not stable across learning situations and school subjects” (p. 77). In addition, the research on student engagement is often built upon a constructivist view of knowledge that education should aid students in constructing their own knowledge (Krause & Coates, 2008; Zepke & Leach, 2010). From the constructivist perspective, student engagement depends not only on the students themselves, but also on educational institutions creating the conditions that “stimulate and
encourage student involvement” (Krause & Coates, 2008, p. 493).

Kuh et al. (2007b) proposed a framework, presented in Figure 1, for student success in higher education that has student engagement at its core.

**Figure 1**

*Kuh, Kinzie, Buckley, Bridges, and Hayek’s (2007b) Framework, “What Matters to Student Success”*


In this framework, the pathway to student success begins with the effects of students’ pre-college experiences, including their family background, precollege academic preparation, enrollment
choices, motivation to learn, and family and peer support. Financial aid, policies, and remediation courses are viewed as mediating conditions that students must successfully transition through in order to begin a college experience. In the center of the framework is the college experience with student engagement being represented as the intersection of (a) student behaviors, such as study habits, involvement with peers and faculty, and time on task, and (b) institutional conditions, such as resources, educational policies, academic support, and the campus environment (Kuh et al., 2007b). Reflecting upon the incorporation of motivation within engagement theories, Kuh et al. (2007b) identified motivation in their framework within the concept of student behaviors (i.e., behavioral engagement); this contrasts with the work of other engagement researchers (Cleary & Zimmerman, 2012; Fredricks et al., 2004) who incorporated motivation within the concept of cognitive engagement. Kuh et al. proposed that student engagement represents two “critical features”: (a) “the amount of time and effort students put into their studies and other educationally purposeful activities” and (b) “how the institution deploys its resources and organizes the curriculum, other learning opportunities, and support services to induce students to participate in activities” that lead to desired educational outcomes (p. 44). This conceptualization of engagement appears to draw mostly upon Astin’s (1984) theory of student involvement and the behavioral dimension of engagement from Fredricks et al.’s (2004) tripartite model. The final portions of the pathway are student learning gains, grades and graduation, leading to post-college outcomes.

In a review and critique of the literature on student engagement, Kahu (2013) identified four dominant research perspectives found across the literature: (a) behavioral, which focuses on student behavior and effective teaching practice; (b) psychological, which focuses on the internal psychosocial processes and state of engagement, (c) sociocultural, which emphasizes the social,
cultural, and political contexts influencing engagement, and (d) holistic, which incorporates aspects of the other perspectives. Kahu argued that the psychological perspective views behavior, cognition, and emotion as antecedents for engagement. The holistic perspective, like the psychological perspective, accounts for emotional influences. The behavioral, sociocultural, and holistic perspectives do not account for antecedents, which Kahu argued is a major weakness in conceptualizing engagement. Each of the four perspectives provide useful insights into student engagement, but alone do not provide a complete conceptualization. Kahu proposed that, in conceptualizing student engagement,

   a clearer distinction would be to recognise that what is considered to be the process [of engagement] is not engagement, instead it is a cluster of factors that influence student engagement (usually the more immediate institutional factors), whereas the outcome is student engagement – an individual psychological state with the three dimensions discussed earlier of affect, cognition and behavior. (p. 764)

Kahu argued that the major limitation of current theories was the lack of distinction between the antecedents, consequences, and state of engagement, and proposed a new framework for student engagement presented in Figure 2.
Kahu’s framework expanded upon Fredricks et al.’s (2004) tripartite model of engagement. It not only incorporated behavioral, emotional, and cognitive engagement, it also clearly specified the relationships of the structural and psychosocial antecedents to engagement, and its proximal and distal consequences. The relationships between student engagement and its immediate psychosocial antecedents (i.e., university, relationships, and student variables) and proximal consequences (i.e., academic and social outcomes) are bi-directional to depict how “engagement breeds engagement” (Kahu, 2013, p. 767). In addition, the framework is embedded within the

sociocultural context of the student and their educational environment. The framework provides a clear theoretical conceptualization of the complexity of student engagement.

In a refinement of Kahu’s (2013) framework of student engagement, Kahu and Nelson (2017) incorporated transition theory and cultural studies into Kahu’s framework and proposed that “individual student engagement occurs dynamically within an educational interface at the intersection of the student and their characteristics and background, and the institution and its practices” and contended that the educational interface “offers a cogent explanation for the dynamic, complex and individual nature of students’ psychosocial learning experiences” (p. 2). Kahu and Nelson’s conceptualization (see Figure 3) of the dynamic nature of engagement aligns with Boekaerts’ (2016) and Cleary and Zimmerman’s (2012) assertions that engagement is malleable and context specific.
Figure 3

Kahu and Nelson’s (2017) Refined Conceptual Framework of Student Engagement

Incorporating the Educational Interface


Transition theory is focused on understanding why many students withdraw or fail during their first year in college and “highlights that alignment, or misalignment, between student and institution, is important for success” (Kahu & Nelson, 2017, p. 5). Kahu and Nelson argued that transition theory can also be utilized to understand attrition and the role of institutions in student success throughout the college experience, not only during the first year. Kahu and Nelson adapted Nakata’s (2007) idea of a “cultural interface” in developing their own concept of the educational interface that places importance on student agency, and on valuing and drawing upon multiple cultural ways of being, rather than taking a deficit stance. In addition, Kahu and Nelson
shifted away from the idea of the temporary state of transition and argued that an educational interface is a psychosocial space where the “student is in a set of relationships within multiple educational settings and their sense of self is dynamic and fluctuating, varying according to the situation being experienced” (p. 6).

Kahu and Nelson (2017) identified four psychosocial constructs as mediating mechanisms (i.e., academic self-efficacy, emotions, belonging, and well-being) that help to understand how students experience the educational interface. According to the authors, either the student and institutional factors (depicted as structural and psychosocial influences within the model) “can interact directly to influence student engagement” or engagement can be mediated through the psychosocial mediating mechanisms (Kahu & Nelson, 2017, p. 7). This refined framework provides insight into “how and where the interactions between institutional and student factors occur” that impact student engagement (p. 2).

**Theoretical approach of study.**

Based upon the conceptualization of engagement utilized in the NSSE, there are two theoretical approaches that best aligned with this study: (a) Fredricks et al.’s (2004) tripartite model of student engagement and (b) Kuh et al.’s (2007b) framework of student success. While the NSSE does not incorporate all three dimensions (behavioral, cognitive, and emotional) of engagement from Fredricks et al.’s tripartite model, it does approach engagement as a multidimensional construct which is the core of the theoretical underpinning of the tripartite model. The NSSE items contribute to an understanding of both the behavioral and cognitive dimensions of the multidimensional engagement construct.

As the NSSE is a survey instrument, there are feasibility constraints as to how many facets of a complex construct you can include in one measurement instrument. Fredricks et al.
themselves wrote, “Essentially, there is a tension between conceptual clarity and practical reality…. it may not be feasible to examine independently each of the ideas included within the types of engagement because of the large number of survey questions that would be required and the time constraints on administering surveys” (p. 84). Acknowledging these constraints, Kuh et al.’s framework of student success was a good complement to the tripartite model within my focus on the NSSE. Fredricks et al. argued that “considering engagement as a multidimensional construct argues for examining antecedents and consequences of behavior, emotion, and cognition simultaneously” (p. 61). Kuh et al.’s framework for student success incorporates antecedents and outcomes of student success with student engagement (including student and institutional factors) at the center. In addition, the creators of the framework for student success include George Kuh, one of the original NSSE design team members, and Jillian Kinzie who is the Associate Director of the NSSE Institute at Indiana University’s Center for Postsecondary Research. With these two influential NSSE scholars having developed the framework and its alignment with the NSSE design, it was a clear fit for the study.

**Defining Student Engagement for the College-Age Population**

Fredricks et al. (2004) argued that the engagement literature is “marked by duplication of concepts and lack of differentiation in definitions across various types of engagement” (p. 65). For example, the concept of effort is often included as part of the definition of both behavioral and cognitive engagement. Other researchers, such as Azevedo (2015) and Boekaerts (2016), have also reached the same conclusion that there is broad lack of consensus on what engagement is and how it should be measured. While there is a lack of consensus on the definition and measurement of engagement, research has identified several different conceptualizations or dimensions of engagement that focus on behavioral, emotional, cognitive, and agentic aspects of
engagement (Fredricks & McColskey, 2012; Sinatra et al., 2015; Trowler, 2010). Different definitions of engagement have been developed based on these different dimensions.

Engagement has been defined as the extent to which students participate in educationally effective practices that are linked to measurable, desired outcomes (Krause & Coates, 2008; Kuh et al., 2007a). Engagement has also been defined as “the process whereby institutions and sector bodies make deliberate attempts to involve and empower students in the process of shaping the learning experience” (Higher Education Funding Council for England, 2008, p. 2). Kuh et al. (2008) included both the role of student and institution in their definition of engagement as “the time and effort students devote to activities that are empirically linked to desired outcomes of college and what institutions do to induce students to participate in these activities (Kuh, 2001)” (p. 542).

For Krause et al. (2005), engagement “refers to the time, energy and resources students devote to activities designed to enhance their learning at university” (p. 31). On the National Survey of Student Engagement (NSSE), student engagement is defined as, “the amount of time and effort students put into their studies and other educationally purposeful activities” (The Center for Postsecondary Research, 2017, para. 1). Within its definition of engagement, the NSSE also emphasizes how higher education institutions allocate resources and organize learning opportunities that encourage students’ participation in the aforementioned educationally purposeful activities (Kuh, 2001a; Wolf-Wendel et al., 2009). The definitions just described fall mainly within what Fredricks et al. (2004) termed the behavioral perspective of engagement.

Student engagement from the behavioral perspective is the "time and effort students devote to educationally purposeful activities" (Radloff & Coates, 2010, p. 1). The behavioral perspective is the most predominant across the literature; however, other perspectives include
more “internal” or psychological aspects of engagement. According to Trowler (2010), engagement "requires feelings and sensemaking as well as activity….Acting without feeling engaged is just involvement or even compliance; feeling engaged without acting is dissociation" (p. 5). Emotional engagement focuses on the affective reactions of students including (a) reactions to peers and faculty, (b) feelings of being valued, (c) interest, (d) enjoyment, and (e) sense of belonging (Fredricks & McColskey, 2012; Trowler, 2010). Sinatra et al. (2015) argued that cognitive engagement is the most difficult dimension to define due to a lack of consensus among scholars of how to operationalize this dimension. Cognitive engagement is often defined as psychological investment, which relates to students’ investment in learning and the learning process (Fredricks & McColskey, 2012; Sinatra et al., 2015; Trowler, 2010). “A student becomes psychologically invested when she or he expends cognitive effort in order to understand, goes beyond the requirement of the activity, uses flexible problem solving, and chooses challenging tasks” (Sinatra et al., 2015, p. 3). Agentic engagement is a more recent addition to the conceptualizations of engagement (Reeve & Tseng, 2011). “This form of engagement occurs when a student actively contributes to the flow of instruction” (Sinatra et al., 2015, p. 3).

The way that these dimensions contribute to the construct of student engagement has also been conceptualized in various ways. Pekrun and Linnenbrink-Garcia (2012) focused on emotions within their multidimensional conceptualization of student engagement and argued that emotions are a precursor to all other components of engagement. These researchers “regard engagement as a mediator between students’ emotions and their achievement” (Pekrun & Linnenbrink-Garcia, 2012, p. 264). In contrast, Kahu (2013) and Kahu and Nelson (2017) consider engagement to be an individual student’s internal psychosocial state comprised of their behavioral, emotional, and cognitive connection to their learning. The definition proposed by
Kahu and Nelson can be classified as a holistic approach. Similarly, Zepke (2015) critiqued the behavioral perspective of engagement as too narrow and proposed that a better approach to understanding student engagement was to utilize a holistic approach that incorporates a sociocultural ecological perspective. Yet another approach is to focus on microlevel aspects of engagement in individual students and define engagement as "the quantity and quality of mental resources directed at an object and the emotions and behaviors entailed" (Miller, 2015, p. 31).

As D'Mello et al. (2017) stated,

engagement has emerged as a broad and complex construct pertaining to diverse aspects of the educational experience (e.g., showing up, completing homework, feelings of belongingness, graduating) and across multiple time scales (e.g., momentary affective episodes, stable dispositions such as general disengagement with school, and life altering outcomes like dropping out of school). (p. 105)

The behavioral, emotional, and cognitive dimensions of engagement all have motivational and self-regulatory aspects within them, which makes the differentiation of these dimensions difficult from a measurement standpoint (Sinatra et al., 2015). In order to address the issues with defining and measuring engagement, Sinatra et al. (2015) recommended thinking of the measurement of engagement as a continuum from person-oriented to context-oriented. Sinatra et al. proposed that, "engagement will be operationalized more effectively if researchers state the dimension of engagement they are exploring, explicitly acknowledge that overlap likely exists with other dimensions, and state the grain size of engagement they are measuring" (p. 8). From this perspective, the measurement of student engagement can be viewed along a microlevel (person-oriented) to a macrolevel (context-oriented) continuum.
Measuring Student Engagement for the College-Age Population

The importance of engagement has led researchers and university administrators to seek out measures of engagement that will provide information for improving undergraduate student success and that provide a means of assessing educational quality (Campbell & Cabrera, 2011; Miller & Malandra, 2006). Sinatra et al. (2015) proposed that researchers “must decide if they are defining engagement as behavioral, cognitive, or emotional, or some combination of multiple dimensions" and this definition should drive their selection of measurement instruments, versus letting the selection of an instrument determine the conceptualization of engagement (p. 7). In a recent review of engagement research within technology-mediated learning environments, Henrie et al. (2015) reported a troubling finding that authors often fail to include any statement of how student engagement is defined within their research. With a clear definition of engagement, the grain size of engagement measurement that best aligns with the intended research can be determined. For example, the researcher “could use physiological measures or eye tracking to measure cognitive engagement, but this could also be accomplished through discourse analysis or classroom observations....this will depend on the researchers’ choice of theoretical framework and research question of interest" (Sinatra et al., 2015, p. 7). Sinatra et al.’s continuum of engagement provided three main reference points for conceptualizing the grain size of engagement measurement and where different theoretical perspectives fall along the continuum: (a) person-oriented, (b) person-in-context, and (c) context-oriented.

**Person-Oriented: End of continuum.**

The finest grain size for measuring student engagement is at the microlevel or person-oriented end of the engagement measurement continuum. In person-oriented engagement research, "engagement is characterized theoretically by the cognitive, emotional, or motivational
engagement of the individual learner” and "the main unit of analysis is typically the individual"
(Sinatra et al., 2015, p. 8). At this microlevel of measurement, the theoretical definition of
engagement focuses on the mental processes of individual students. When examining
engagement at this microlevel, scholars may utilize physiological or psychophysiological
indicators such as eye-tracking or heart rate. Other examples of person-oriented measurement
approaches include reading response times and strategic behaviors.

D'Mello et al. (2017) developed a measurement approach, which they called advanced,
analytic, and automated (AAA) measurement, focused on measuring the “state (not trait) of
engagement across microlevel time scales ranging from seconds to minutes” using advanced
computer technology (p. 106). Their research highlighted the mind-body link that “observable
bodily responses can be used to infer unobservable mental states…associated with engagement
(e.g., concentration, interest) from machine-readable signals and from aspects of the
environmental context” (D'Mello et al., 2017, p. 107). The AAA measurement approach has
been utilized with various measures to record signals while college-age students complete
learning activities, including eye gaze data in combination with key stroke analysis (Bixler &
D'Mello, 2013), a machine-learned model based on eye gaze data to detect mind wandering
during computerized reading (Faber et al., 2017), and video recordings for analysis of facial
expressions (Whitehill et al., 2014).

The advantage of this type of AAA measure is that it can measure cognitive engagement
automatically and without disrupting the learning process. This AAA approach is particularly
well-suited for online learning environments. However, with microlevel measures it is difficult to
disentangle the different dimensions of engagement. Thus, one of the limitations of this type of
measure is that results are often difficult to interpret because engagement has to be isolated from
other factors. Triangulation with other measures, such as self-reports of engagement, can aid in the interpretation of microlevel engagement data and compensate for the limitations of these person-oriented measures. Research using these different types of engagement measures can fall along Sinatra et al.’s (2015) engagement continuum between the mid- and end-points depending upon the theoretical framework and grain size of measurement utilized.

**Person-in-Context: Middle of continuum.**

Moving along to the middle of the engagement measurement continuum, scholars utilizing a person-in-context theoretical framework aim to understand an individual within a particular context, with the focus of analysis being the interactions themselves (Sinatra et al., 2015). The goal of person-in-context engagement research is to "assess how student(s) interact with others and with dimensions of their environment (i.e., classroom instruction, technology) to produce a particular type, level, or form of engagement" (Sinatra et al., 2015, p. 9). Researchers analyze students’ cognition, emotions, or behaviors in response to a particular environment. Methods that can be utilized for the person-in-context approach to engagement include experience sampling, observations of interactions, and self-report of engagement and student experiences. Surveys of student engagement, such as the NSSE, usually fall within the person-in-context framework as they focus on how students are engaging within a particular institutional context.

Greene et al. (2003) utilized the person-in-context approach in their study of chemical engineering college students to investigate the self-efficacy, engagement, and achievement of students in two different instructional contexts: technology-based instruction and traditional instruction. These researchers utilized a combination of self-report and interview methodology. Research utilizing the person-in-context approach has shown that the subject domain (i.e., math,
history, etc.) as well as the type of task being performed matter when measuring engagement. Engagement measures can function differently depending on the subject domain and task type being studied (Greene, 2015). In order to gather more immediate data on engagement within context, as opposed to recalled self-reports, experience sampling techniques can also be used with mobile technologies, such as iPads and smartphones, to gather data on students’ engagement at multiple time points throughout a study period (Greene, 2015; Sinatra et al., 2015). In addition, the experience sampling method can be used to capture variance in student engagement over time (Henrie et al., 2015). One limitation of the experience sampling method, however, is that it often requires significant effort from students to complete and may be intrusive in the learning process.

The person-in-context approach to engagement measurement allows the researcher to explore the ways that a student’s engagement is influenced by social interaction and particular educational contexts. Hilpert and Husman (2017) utilized a combination of observations of faculty teaching practices and students’ self-report on their engagement levels to examine the impact of a year-long faculty professional development program on the engagement levels of undergraduate students. While observation can provide rich data, a limitation of observational measures is that the results are often not generalizable and, in addition, this method requires significant time and training of observers, which limits its scalability (i.e., the ease with which this measure can be implemented at a larger scale). In a critique of the measurement of student engagement, Azevedo (2015) argued that measurement of engagement based on self-report, “can hinder our understanding of the deployment of engagement processes…. the problem is that self-reports are based on students’ perceptions of how one would or did enact certain processes, and these perceptions often do not align with what actually occurs during learning (Zimmerman,
Based on this potential limitation of self-reports, Azevedo proposed that researchers should utilize multiple methodologies to capture engagement processes including the microlevel person-oriented measures, person-in-context measures, and macrolevel context-oriented measures.

**Context-Oriented: End of continuum.**

The largest grain size for measuring student engagement is the macrolevel or context-oriented end of the engagement measurement continuum. In context-oriented engagement research, "researchers focus on capturing the characteristics of the classroom, school, community, or culture that afford or impede engagement" (Sinatra et al., 2015, p. 9). Macrolevel research examines engagement from a sociocultural and situated cognition perspective (Sinatra et al., 2015). Engagement from this perspective is a dynamic and reciprocal process that connects collective and individual engagement, and is “conceptualized as meaningful changes in participation” (Ryu & Lombardi, 2015, p. 71). Examples of context-oriented measures include discourse analysis, observations, and social network analysis.

Ryu and Lombardi (2015) argued that to be consistent with the macrolevel sociocultural perspective, researchers must use “methods that go beyond individual measurement by characterizing and analyzing engagement as changes in participation that occur when students engage in social and relevant disciplinary practices (e.g., science learning in classroom communities)” (p. 70). In this perspective, cognition is mediated through social activity and individual engagement involves taking on new roles and responsibilities that aid in the development of an individual’s identity and agency (Ryu & Lombardi, 2015). The authors proposed that an effective means for linking collective and individual engagement is through the use of a mixed method approach that combines critical discourse analysis and social network
analysis. Critical discourse analysis examines language use and the relationship of language use with social interactions, as well as implications of status and power (Ryu & Lombardi, 2015). Social network analysis aims to characterize the relationships between individuals and the patterns of interaction among them; it allows researchers to “visualize engagement by tracing the shape of and changes in participation over time” (Ryu & Lombardi, 2015, p. 76).

Casimiro (2016) used discourse analysis of an online discussion forum in a graduate level course using two units of analysis, the group and individual students, to investigate the conditions that support engagement in online classes. While the study was very small scale in size, the use of discourse analysis to identify contextual factors that afford or impede student engagement provides a good example of research that can be done along the context-oriented end of the engagement continuum. In another example of a context-oriented approach, Zhu (2006) utilized social network analysis combined with qualitative content analysis to examine undergraduate and graduate students’ cognitive engagement during online discussions in order to identify factors that may cause differences in interaction and levels of engagement. These studies highlight that the utilization of context-oriented approaches, such as critical discourse analysis, observations, and social network analysis, provide researchers with a means of exploring the dynamic relationship of collective and individual engagement, and changes in meaningful participation over time.

Validation Studies of Student Engagement Measures

The previous discussion has shown that student engagement measures are diverse and can span the continuum from person-oriented to context-oriented measures. This study was focused on the NSSE, a self-report instrument, which I would classify as a person-in-context measure using Sinatra et al.’s (2015) continuum of engagement measurement. As this study utilized data
from a self-report instrument and most large scale studies of student engagement with college-aged populations have used self-report instruments, I focused my comparison of validation studies on three influential and large scale self-report engagement measures used with college-age students (Klemenčič & Chirikov, 2015; Maskell & Collins, 2017): (a) the NSSE, (b) the Student Experience in the Research University (SERU) Survey, and (c) the National Student Survey. The NSSE has been adapted into other NSSE-based national surveys, such as the Australasian Survey of Student Engagement (AUSSE) and the United Kingdom Engagement Survey (UKES); however, because these are based on the NSSE, I did not include them in the validation study comparison.

*The National Survey of Student Engagement.*

The construct of student engagement in higher education has ranging conceptualizations from a narrower focus on singular aspects of the construct, such as behavioral or emotional, to more holistic approaches, such as the framework proposed by Kahu (2013). The construct of student engagement within the NSSE falls more along the narrow end of the spectrum, as opposed to the holistic approach. The NSSE was created as a project in the context of the 1990s when there was growing dissatisfaction with college rankings and measures of institutional quality in the United States (Kahu, 2013; Kuh, 2009). The original design team for the NSSE was comprised of influential scholars in the field of student engagement including Alexander Astin, Arthur Chickering, and George Kuh (The Center for Postsecondary Research, 2016). The design team developed the NSSE largely based on Chickering and Gamson's (1987) “Seven Principles of Good Practice in Undergraduate Education,” which drew upon the behavioral perspective of engagement. The seven principles were:

1. Encourages contacts between students and faculty.
2. Develops reciprocity and cooperation among students.

3. Uses active learning techniques.

4. Gives prompt feedback

5. Emphasizes time on task

6. Communicates high expectations.

7. Respects diverse talents and ways of learning. (Chickering & Gamson, 1987, p. 3)

In addition to the work of Chickering and Gamson, the NSSE design team also drew upon other seminal research including Pace's (1980) quality of effort, Astin's (1984) theory of involvement, Tinto's (1987, 1993) theory of departure, and Pascarella's (1985) college outcomes. Most of the original items for the NSSE came from other well-researched instruments; with two-thirds coming from the College Students Experience Questionnaire (CSEQ). The survey is administered to freshman and senior undergraduate students in the spring semester of the academic year of its administration. The first national administration of the NSSE was in 2000 and the number of participating institutions has steadily grown since then (Kuh, 2009). In 2015, over 315,000 undergraduate freshman and seniors from 541 U.S. institutions responded to the survey (The Center for Postsecondary Research, 2015). The NSSE is intended to be used by institutions to “identify aspects of the undergraduate experience inside and outside the classroom that can be improved through changes in policies and practices more consistent with good practices in undergraduate education” (The Center for Postsecondary Research, 2017, "How are survey results used?" section).

The NSSE, which is currently used in the U.S. and Canada, was originally designed with five major engagement scales, and was later modified by the addition of a sixth scale into the Australasian Survey of Student Engagement (AUSSE), used in Australia and New Zealand:
1. Level of Academic Challenge (extent to which activities and conditions challenge students to learn);
2. Active and Collaborative Learning (students’ efforts to actively and collaboratively construct their knowledge);
3. Student-Faculty Interaction (level and nature of students’ interaction with faculty);
4. Enriching Educational Experiences (participation in broadening educational activities);
5. Supportive Campus Environment (feelings of legitimation within the university community);
6. Work-integrated learning (integration of employment-focused work experience into study). *This factor is only present in the AUSSE.* (Trowler, 2010)

In 2009, the Center for Postsecondary Research began a multi-year project for updating the NSSE which resulted in a change from the five NSSE benchmarks to ten “Engagement Indicators” that fall within themes, as presented in Table 1 (McCormick et al., 2013). In contrast to the five NSSE benchmarks, the ten engagement indicators were designed by the NSSE staff to be unidimensional constructs (The Center for Postsecondary Research, n.d.). The revised NSSE has 47 items that comprise the ten engagement indicators, and has additional items for six high-impact practices and demographics (see Appendix A for the 2015 NSSE instrument and Appendix B for the relation of items to engagement indicators). Engagement indicator scores are intended to provide information about distinct components of student engagement, for institutional comparisons, and for comparisons over time. The scores are calculated by (1) recoding the engagement indicator item responses on a 60-point scale, (2) averaging together the recoded values for each engagement indicator, and (3) weighting the averages based on gender.
The NSSE was primarily based on the behavioral perspective of student engagement. The items on the NSSE focus on institutional actions and requirements, student behavior, and students’ perceptions about the quality of their college experiences (The Center for Postsecondary Research, 2016). Kahu (2013) argued that the NSSE also does incorporate the cognitive aspect, but does not include the emotional aspect of engagement. The underlying assumptions and student engagement construct definition of the NSSE, therefore, are focused on the behavioral and cognitive aspects of engagement.

Multiple studies using national and multi-institution data have found evidence that the NSSE provided valid and reliable data for institution and group level assessment and research of
student engagement (Kuh, 2001b; Miller et al., 2016; Ouimet et al., 2004; Pascarella et al., 2010; Pike, 2006, 2013a; Zilvinskis et al., 2017). These studies have shown that the NSSE items and scales are theoretically and empirically derived, and are psychometrically sound. As the revised NSSE is relatively new, there are relatively few published validation studies on the new 2013 version. Using national data from the NSSE in 2008, Pike (2013a) found that the NSSE benchmarks “provided dependable means for 50 or more students and were significantly related to important institutional outcomes such as retention and graduation rates” (p. 149). Looking at the revised 2013 version of the NSSE, Miller et al. (2016) analyzed national data using exploratory and confirmatory factor analysis and found evidence for the validity of the internal structure of the revised NSSE’s ten engagement indicators organized by four themes. Zilvinskis et al. (2017) used canonical correlation analysis to examine the relationship between the NSSE and self-reported learning gains using national data for first-year students from both the 2011 and the revised 2013 versions of the survey. Results of the study provided convergent evidence for both NSSE versions with the student engagement measures being significantly and positively related to students’ self-reported learning gains. In addition, results indicated that only the new 2013 version displayed discriminant evidence with the engagement scales differentiating among the self-reported learning outcomes.

While there are numerous studies indicating the psychometric soundness of the NSSE, there are also numerous studies with contrary findings, thus making the NSSE a somewhat controversial instrument. In countering the debate surrounding the NSSE, McCormick and McClenny (2012) proposed that “student engagement surveys were designed for consequential validity—that is, to produce data that are meaningful and actionable—in other words, information that is good enough to be useful in decision-making” (p. 331). While providing
information that is “good enough” may be acceptable in some circumstances, research has identified limitations in the NSSE and its applicability. Porter (2011) focused on the NSSE in his critique of survey instruments; he challenged students’ ability to recall information reliably given an extended time frame for recall, such as an academic year. This general critique of self-report instruments can be applied to the SERU and NSS as well. In a rebuttal to Porter’s critique, McCormick and McClenny (2012) argued that the validation of an instrument must be based on the intended uses and proposed interpretations of data, and that Porter’s position neglects the way that NSSE results are used, which is to make comparisons between groups of students. McCormick and McClenny (2012) argued that “analyses of data from NSSE…suggest that discrepancies in individuals’ and groups’ uses of…response options do not meaningfully limit how the data are typically used (Cole & Korkmaz, 2011; Nelson Laird, Korkmaz, & Chen, 2009)” (p. 315-316). The authors also countered that Porter’s focus on individual students’ accurate recall disregards the fact that for the NSSE, “What matters is not the precise number of papers written but the fact that certain groups of students write more than others” (McCormick & McClenny, 2012, p. 315).

Recent studies (Campbell & Cabrera, 2011; Gordon et al., 2008; LaNasa et al., 2009; Tendhar et al., 2013) of institutional level NSSE data for single, public, research-intensive institutions prior to its 2013 revision have found that the NSSE benchmarks (latent construct) model was a poor representation for their data, calling into question the validity evidence based on the internal structure of the NSSE. In addition, Gordon et al. (2008) questioned the predictive (i.e., test-criterion) validity evidence of the NSSE in relating the benchmarks to outcomes, such as grade point average and student retention, based on a dearth of research in this area and their single institution study that found only modest predictive power of NSSE benchmarks in
explaining student outcomes. In a study examining the linkages between the NSSE benchmarks and student success outcomes at 14 institutions, Carini et al. (2006) found very weak associations, also calling into question the evidence of test-criterion relationships for the NSSE. The results differed across institutions and there were stronger linkages for students with lower ability levels. Based upon the poor fit of the NSSE benchmarks to their institutional data, LaNasa et al. (2009) found that "the strongest implication is a cautionary one: It is incumbent on institutions to fully explore their own data, especially when using the data for comparative purposes. To fully understand NSSE results, item level inspection is required" (p.330).

There is a dearth of research, however, exploring whether the uses and interpretations of the revised NSSE scores are psychometrically sound. Fosnacht and Gonyea (2018) examined the generalizability of the NSSE 2.0 engagement indicators with national data and found that the engagement indicators can reliably be generalized from small samples to a larger population of students; however, they also found that only some of the engagement indicators can dependably be generalized to higher-order constructs using small samples. Two validation studies of NSSE 2.0 (Le, 2019; Winkler, 2020) on single institution’s data raised concerns about the reliability and validity of the NSSE’s ten engagement indicators and four themes on their data sets. Winkler (2020) examined the NSSE 2.0’s ten engagement indicator score utility using classical test theory and generalizability theory methods; the study found that the engagement indicator scores were reliable at the individual student level, but “were not sufficiently reliable nor substantively meaningful to warrant use for institutional decision-making” at the college level (p. iii). Le (2019) utilized CFA to study the fit of the NSSE 2.0 engagement indicators organized into four themes model for transfer students and found mixed results of model fit across the four themes and transfer student subgroups, with only one theme, Academic Challenge, demonstrating good
fit across the examined groups. Only one published study (Carle et al., 2009) was found that utilized item response theory to analyze the NSSE. This study utilized NSSE data prior to 2013 when the instrument underwent a significant revision. Research has shown that changes to the content of a measurement instrument can have significant influences on the psychometric properties of the instrument, and therefore when an instrument undergoes major revisions, its psychometric properties must be reexamined to ensure that the uses and interpretation of scores are psychometrically appropriate (Furr, 2011).

Student engagement as measured by the NSSE does not fully encompass the broad scope of the construct as proposed by researchers such as Kahu (2013). Leach (2016) concluded that because the AUSSE, which was developed from the NSSE, was based on constructivism and constructivist educational practices, disciplines that utilized an instructivist approach were at a disadvantage on the instrument. Leach argued that while the AUSSE was based on the assumption of constructivism, many items were written based on "a behaviorist approach to student engagement," and that the instrument does not measure the emotional aspect of engagement, nor does it measure the sociocultural contexts that impact engagement; this argument also applies to the NSSE (p. 783). The narrow definition of student engagement found within the NSSE and AUSSE may be seen as a potential limitation, especially in comparison to the expanded conception of engagement as found in Kahu’s (2013) framework. In addition, research on the NSSE has indicated that there are significant differences in levels of student engagement based on class standing of students, the academic discipline of students, and age of students, traditional versus nontraditional (Kuh, 2001b; Pike, 2013b; Popkess & McDaniel, 2011; Price & Baker, 2012). Findings suggest that institutions who utilize the NSSE should be mindful of carefully investigating the ways in which the tool measures engagement across their
diverse student populations.

**The Student Experience in the Research University Survey.**

The Student Experience in the Research University (SERU) survey was developed as a measure of student experiences, including student engagement, at research intensive universities. The survey was developed in the early 2000s to be used across the University of California (UC) system and focused on academic and civic engagement of undergraduate students (Center for Studies in Higher Education, 2017). The project has since expanded to include a consortium of research intensive universities across the U.S. and internationally, including institutions in Brazil, China, Europe, and South Africa. Institutions voluntarily participate in the survey and results are intended to be used for institutional comparison and improvement. There are both undergraduate and graduate versions of the SERU. The undergraduate SERU is a census and is administered to all undergraduate students in the target population, unlike the NSSE which is administered to a random sample of first-year and final-year students. The content of the SERU is focused on student engagement at research universities in three areas: teaching and learning, research, and civic service (Klemenčič & Chirikov, 2015, p. 366).

Multiple studies have found supporting validity and reliability evidence for the SERU. Chatman (2009) reported reliability evidence and validity evidence based on instrument content and internal structure for the SERU utilizing a combination of exploratory factor analysis and collaborative expert judgement from a research team. Chatman (2011) found that validation results of the SERU factor structure through exploratory factor analysis using random samples of SERU data from research universities across the U.S. have been relatively similar across four independent factor analyses in 2006, 2008, 2009, and 2011. The differences in results were most attributed to changes in the survey structure, such as moving, adding, or deleting items. In
another study, Douglass et al. (2012) found validity evidence based on test-criterion relationships of the SERU for assessing students’ learning gains utilizing the 2008 SERU data and actual grade point average data from UC seniors. The authors concluded that, “Although self-reported gains are sometimes regarded as having dubious validity compared to so-called "direct measures" of student learning, the analysis of this study reveals the SERU survey design has many advantages, especially in large, complex institutional settings” (Douglass et al., 2012, p. 317). Hernandez et al. (2013) examined validity evidence based on the internal structure of the SERU for Latino students at one research intensive university through exploratory factor analysis. Their results supported the proposed SERU factor structure. They also found that there were cultural nuances in the SERU subfactors that emphasized the “role of culture, sense of agency, initiative-taking, self-competency, and self-efficacy” (Hernandez et al., 2013, p. 2).

Similarly to NSSE research, scholars have found differences in engagement levels as measured by the SERU based on student demographics, academic ability level, and academic field of study (Douglass et al., 2012; Hernandez et al., 2013; Soria & Stebleton, 2012). Soria and Stebleton (2012) analyzed SERU data from one research intensive university and found differences in student engagement levels for first-generation and non-first-generation freshman undergraduate students, with first-generation students having lower levels of engagement and lower retention rates. While Hernandez et al. (2013) examined validity evidence based on the internal structure of the SERU using factor analysis, there is a dearth of research that has carried out item-level analysis of student engagement surveys, such as the SERU and NSSE, to establish validity evidence for their use with diverse cultural groups. Many of the limitations of the NSSE, also apply to the SERU, such as Porter’s (2011) critique of students’ recall, the need to be cognizant of differences in student engagement across diverse groups of students, and the
emphasis on a narrower behavioral perspective of engagement. It also is important to keep in mind that the target population of the NSSE and SERU are not the same as the SERU focuses only on students in research universities, while the target population of the NSSE is any four-year higher education institution.

**The National Student Survey.**

The National Student Survey (NSS) was developed in 2005 by the Higher Education Funding Council for England on behalf of the United Kingdom (U.K.) higher education funding bodies in order to increase transparency in higher education institutions across the U.K., provide prospective students with information for making informed choices, and to be used for institutional improvement (Maskell & Collins, 2017). The NSS is administered to all final-year undergraduate students in England, Scotland, Wales, and Northern Ireland at publicly funded higher education institutions, as well as at alternative providers in England. The NSS is comprised of 27 core items; making it a much shorter instrument than the NSSE, and somewhat shorter than the SERU. As opposed to the NSSE and SERU whose primary intended use is institutional improvement, the NSS’ primary intended use is accountability and informing student choice. The NSS data is publicly available on a website (www.unistats.direct.gov.uk) where stakeholders, such as prospective students, can search through the data in a very user friendly way and create custom comparisons of programs across the U.K. to see how institutions and programs compare with each other based on NSS data.

The NSS is “primarily concerned with the assessment of student course experience and seek to capture the various facets of the student learning process” (Klemenčič & Chirikov, 2015, p. 366). In the U.K., they call academic degree programs “courses” so the difference in terminology could confuse a U.S. reader into thinking their items are asking about one U.S. style
course offered within a degree program; however, the survey is intended to measure student learning and engagement in a degree program. Ramsden and Callender (2014) proposed that the concept of engagement in the NSS is “seen to be about how teaching and course organisation influence what students do and how they learn” and also “about learning and development gains themselves” (p. 28-29).

Several studies have provided validity and reliability evidence for the NSS; however, there is much less published research on the NSS than for the NSSE. When the NSS was under development, Richardson et al. (2007) analyzed the 2003 and 2004 NSS pilot data and found evidence of satisfactory levels of internal consistency and internal structure using principal components analysis. After the first two full administrations of the NSS, Marsh and Cheng (2008) examined the dimensionality and multilevel structure of the 2005 and 2006 NSS data. Factor analysis results indicated that a six-factor solution was the best fit, which supported the original proposed structure of the NSS. Using multilevel modeling, the researchers also found that the differences between institutions in the overall satisfaction item were very small (only 2-3% of explained variance) but reliable and stable over time, and that degree program (i.e., subject area) within universities explained more variance than institution. This prompted Marsh and Cheng (2008) to encourage future research to address the question of whether very small (highly reliable) differences between universities were “sufficiently large to help inform the choices of prospective students – a primary purpose of the NSS” (p. 52).

Using multilevel modeling techniques, Surridge (2009) examined differences between groups of students and variations between institutions once student profiles were controlled for in the 2005, 2006, and 2007 NSS survey results. Surridge (2009) found that each of the scales had high internal consistency and there were positive associations between the scales and overall
satisfaction item. Callender et al. (2014) reviewed the existing research on the NSS in addition to collecting feedback from higher education stakeholders. Based on their review and stakeholder feedback, the authors confirmed the validity and reliability evidence for the instrument’s intended uses and score interpretations, while also suggesting significant revisions to the NSS to address issues of the scope of the instrument. Callender et al.’s (2014) findings led to the NSS being significantly revised for its 2017 administration; included in the revisions was the addition of nine new items on student engagement, with a focus on academic challenge, the learning community/collaborative learning, and student voice (Higher Education Funding Council for England et al., 2016).

Some scholars have critiqued the NSS as not including “students’ own contribution to their learning and development” within its conception of student engagement (Klemenčič & Chirikov, 2015, p. 367); however, this limitation was addressed in the 2017 revision. Ramsden and Callender (2014) proposed that while the NSS “effectively identifies strong points and areas of concern regarding the quality of teaching and courses, it says nothing about why some groups of students have more positive experiences than others” (p. 44). Also, as the NSS is only administered to final-year students, it does not include responses from students who did not complete the program in earlier stages, which could be seen as losing valuable information from that subset of students who potentially could be most dissatisfied (Callender et al., 2014). Similarly to the NSSE and SERU, researchers have also found differences in NSS scores by student profiles, such as age, gender, ethnicity, and by academic subject area (Marsh & Cheng, 2008; Surridge, 2009).
Item Response Theory: Unidimensional and Multidimensional Examination

Most of the research investigating the NSSE has utilized principles based on classical test theory (CTT); however, item response theory can provide much richer analyses of measurement precision than can be achieved with CTT (Embretson & Reise, 2000; Sharkness & DeAngelo, 2011). Through a review of the literature, only one study was found that utilized IRT to analyze the NSSE and this study was done prior to the significant revision of the instrument. Carle et al. (2009) utilized IRT to develop new engagement scales with the NSSE items, as opposed to investigating the psychometric properties of the individual items and the original five factor NSSE model.

IRT is a latent trait model-based measurement that allows for the estimation of both item level psychometric properties and the latent ability level of individuals (Embretson & Reise, 2000). An IRT model is a mathematical function that is used to describe the conditional probability of an item response based on latent ability level. The probability of responding in a particular way is seen to be influenced by the person’s underlying latent ability level, as well as by the difficulty level of items (Furr & Bacharach, 2014). In IRT, items also differ in the degree to which they differentiate between respondents based on latent ability level (Furr & Bacharach, 2014). Logistic and normal ogive mathematical functions are commonly used in IRT models, both of which produce similar probabilities and parameter estimates, but logistic functions are more often used because of their computational simplicity (Embretson & Reise, 2000). There are multiple IRT models; the selection of an appropriate model varies depending upon the type of item response categories, number of latent traits, and number of item parameters. For unidimensional scales (i.e., those with one underlying latent trait) that contain Likert-type ordered polytomous response categories, such as is found in the response categories for the
NSSE’s engagement indicator items, Samejima’s (1969) Graded Response Model (GRM) is the recommended IRT model. The GRM is one of the most widely used IRT models developed for ordered categorical responses. When considering whether a sample size is sufficient for IRT analysis, there is not a definitive minimum number of respondents; however, research has indicated that a sample size of at least 500 is recommended for accurate parameter estimates (Embretson & Reise, 2000; Thorpe & Favia, 2012). As the UHM’s 2015 sample for the NSSE included 1,592 respondents, the sample size was more than sufficient for IRT analyses.

The parameterization of GRM considers the lowest score on item $i$ to be 0 and the highest score to be $m_i$. The probability of endorsing $j$ or more categories is assumed to increase monotonically with an increase in the underlying latent trait, $\theta$ (Reckase, 2009). In other words, the probability of responding in a particular way on an item is a function of the distance between an item’s location and person’s location. The expression of the graded response model is:

$$P(X_{is} \geq j|\theta_s,\beta_{ij},\alpha_i) = \frac{e^{\alpha_i(\theta_s-\beta_{ij})}}{1 + e^{\alpha_i(\theta_s-\beta_{ij})}}$$

$X_{is}$ refers to a particular response ($X$) made by respondent $s$ to item $i$; $\beta_i$ is the difficulty (i.e., threshold parameter) of an item; and $\alpha_i$ is the discrimination of an item (Furr & Bacharach, 2014). If the probability of endorsing response category $j$ at a particular theta ($\theta$) level is $P(X_{is} \geq j|\theta_s,\beta_{ij},\alpha_i)$, then the probability that a respondent will endorse category $j$ is:

$$P(X_{is} = j|\theta_s,\beta_{ij},\alpha_i) = P(X_{is} \geq j - 1|\theta_s,\beta_{ij},\alpha_i) - P(X_{is} \geq j|\theta_s,\beta_{ij},\alpha_i)$$

Samejima’s (1969) GRM is based upon the assumption of unidimensionality. However, there are times when more than one underlying trait is considered to affect a respondent’s performance. When an instrument is intended to measure multiple traits which are involved in the response to an item, such as the NSSE 2.0 structure of engagement indicators and higher level themes, multidimensional item response theory (MIRT) models should be considered (Kuo
& Sheng, 2016). Figure 4 illustrates a simple example of unidimensionality and multidimensionality.

**Figure 4**

*Examples of Simple Unidimensional and Multidimensional Models*

MIRT builds upon, and can be considered an extension of, unidimensional IRT. The expression of the MIRT model is:

\[
P(X_{is} \geq j | \theta_s, \beta_{ij}, \alpha_i) = \frac{e^{\sum(a_i(\theta_s - \beta_{ij}))}}{1 + e^{\sum(a_i(\theta_s - \beta_{ij}))}}
\]

Whereas in unidimensional IRT the probability of responding in a particular way on an item is a function of the distance between an item’s location and person’s location; in MIRT, this same probability is on multidimensional latent space (Reckase, 1997).

When unidimensional IRT models are applied to multidimensional data, research has shown that this usually results in biased parameter estimates (Ackerman, 1989; Ansley & Forsyth, 1985; Folk & Green, 1989; Way et al., 1988). Research has investigated the differences in performance of subscale ability estimation for MIRT and separate unidimensional IRT on
each subscale, with results indicating that MIRT provides more accurate estimates since it can account for correlations among subscales (de la Torre, 2008; de la Torre & Patz, 2005; de la Torre & Song, 2009; Sheng & Wikle, 2007; Wang et al., 2004; Yao & Boughton, 2007). In addition, research has indicated that MIRT provides better estimation when there are large numbers of dimensions and when there are a small number of items in each dimension (de la Torre, 2008; de la Torre & Patz, 2005; de la Torre & Song, 2009; Wang et al., 2004); both of these conditions are found within the revised NSSE as there are 10 engagement indicators, with some being comprised of only three items.

Based upon my review of the literature, there have not been any studies published which applied MIRT estimation methods to the NSSE. Until recently, software programs were limited in their capability to conduct MIRT on instruments with polytomous ordered response options and complex construct structures (i.e., large numbers of latent dimensions) (Chalmers, 2012; Kuo & Sheng, 2016). The suggested structure of engagement indicators which fall under higher level themes can be investigated utilizing MIRT in order to gain the most informative and accurate measurement estimations.

**Confirmatory Factor Analysis: First-Order, Bifactor, and Higher-Order Factor Structures**

As with MIRT, it is possible to confirm the latent construct structure of the NSSE using CFA. Factor analysis and MIRT can be considered equivalent methodologies despite having different intended uses and focuses (i.e., defining factors versus investigating interactions of items and individuals), with factor analysis being more prevalently used by researchers (Reckase, 1997). The models illustrated in Figure 1 and Figure 2 are also applicable to CFA. In measurement models with a first-order factor structure, items load onto single factors. In a bifactor model, each item *simultaneously* loads onto (i.e., the item is measuring) one general
(common) latent dimension, as well as loading onto a specific (secondary) latent dimension
(Martelli, 2014; Rijmen, 2010). In contrast to bi-factor models, second-order models have a
hierarchical structure between the general and specific latent constructs: items load onto a
specific dimension, and the specific dimensions in turn load onto (i.e., are influenced by) the
general dimension. Applying this to the NSSE, in a first-order model, NSSE items would load
solely onto an engagement indicator; in a bi-factor model, the items would load simultaneously
onto an engagement indicator and a general trait; and, in a second-order model, the items would
load onto an engagement indicator, which in turn would load onto a theme. In a third-order
model, the second-order factors load onto a higher level overall construct; in the case of the
NSSE, this would indicate that the four NSSE themes load onto a higher-order overall factor,
student engagement.

In a common CFA model, the researcher seeks to test the number of factors and
relationship patterns between factors and items based on an a priori hypothesis, informed by
theory and prior evidence. The CFA model is specified to reflect the hypothesized relationship
between items, factors that represent the latent traits the items are supposed to be measuring, and
measurement error that accounts for all the unique sources of variance that are not explained by
the factors (Long, 1983). The relationship between factor(s) and items (i.e., observed variables)
is expressed through factor loadings, which are generally interpreted as regression coefficients.
The parameters included in all CFA models are the factor loadings, factor variance which
represents the variability of a factor, and unique variance which typically refers to measurement
error. The CFA model can be expressed as: \( X_{is} = \lambda_{id} \theta_{sd} + e_{is} \), where \( \lambda_{id} \) is the factor loading
(weight) for item \( i \) on factor \( d \); \( \theta_{sd} \) is the factor score for respondent \( s \) on factor \( d \), which
corresponds to the latent trait \( \theta_s^{(d)} \) in IRT; and \( e_{is} \) is the estimated unique variance for respondent
s on item \(i\). Either a variance-covariance matrix or a correlation matrix, which is a standardized form of a variance-covariance matrix, are used as input. If using a correlation matrix with polytomous response options, such as is found in the NSSE, a polychoric correlation would be used, which is an estimate of correlation of the normally distributed continuous latent traits based on the observed polytomous responses.

When instruments measure both general and specific latent traits, hierarchical models can be utilized (de la Torre & Song, 2009). Higher-order models, such as the second-order model, involve more general dimensions at a higher level and specific domain traits at a lower order. In a second-order model, each item loads onto one specific domain trait, and these specific domain traits load onto an overall trait, the second-order. The CFA second-order model can be expressed by two equations,

\[ Y = \Lambda_i^{(d)} \eta + \varepsilon \] (equation one) and \[ \eta = \Gamma \xi + \zeta \] (equation two), where \(Y\) is a vector representing the observed variables (i.e. items), \(\Lambda_i^{(d)}\) represents the loadings of the items on the first-order factors, \(\eta\) is a vector representing the first-order factors, \(\Gamma\) is a vector representing the loadings of the first-order factors on the second-order factor, \(\xi\) is a vector representing the second-order factor, \(\zeta\) is a vector representing the unique variance of the first-order factors (i.e. unique variance not shared with the second-order factor), and \(\varepsilon\) represents the residuals (unique) variance (Chen et al., 2006). Equation one represents the measurement model for the observed variables: \(Y = \Lambda_i^{(d)} \eta + \varepsilon\). Equation two represents the structure for each of the first-order factors: \(\eta = \Gamma \xi + \zeta\).

In a bifactor model, there is a general trait that has a direct effect on all of the items, but does not have an effect on the specific domain traits. This contrasts with the second-order model in which the general or overall trait has a direct effect on the specific domain traits, but an indirect effect on all the items through the specific domain traits (see Figure 5).
Figure 5

*Illustration of Second-Order and Bifactor Model Structures*

The CFA bifactor model can be expressed by the equation: 

\[ Y = \Lambda_i^{(d)(G)} \eta^{(d)(G)} + \varepsilon, \]

where \( Y \) represents the observed variables, \( \Lambda_i^{(d)(G)} \) represents the factor loadings of the general (G) and specific domain (d) factors, \( \eta^{(d)(G)} \) is a vector representing the general (G) and specific domain (d) factors, and \( \varepsilon \) represents the unique variance (Chen et al., 2006). The specific domain traits account for the residual variance shared by subsets of items (DeMars, 2013).

The justification for selecting a higher-order or bifactor model is largely based on the theoretical hypothesis of the relationship between latent traits and respondents’ (i.e., college students in the case of the NSSE) performance on the instrument. In a bifactor model, it is assumed that the variability in respondents’ performance is accounted for directly by both the general trait and specific domain traits, whereas in a higher-order model the variability in respondents’ performance is accounted for directly by only the specific domain traits. Higher-order models can be seen as more flexible than bifactor models because they can include more
than one general trait and allow for nonlinear relationships between the higher-order trait and the specific domain traits (de la Torre & Song, 2009).

While CFA and MIRT are formulated differently, research has shown that they are mathematically equivalent (Knol & Berger, 1991; McLeod et al., 2001; Takane & de Leeuw, 1987). In a common factor model, it is assumed that the response variable for item $i$, $X_i$, is governed by a continuous latent variable $\lambda_i$, and the threshold parameter $\tau_i$ which corresponds to the location on a scale that separates (i.e., dichotomizes “1” versus “0”) an item’s response categories. Therefore, the common factor model equation, $X_{is} = \lambda_{id} \theta_{sd} + \epsilon_{is}$, can be rewritten as a normal distribution function $\Phi$ for a “1” response: $P(X_{is} = 1|\theta) = \Phi \left( \frac{\lambda_i \theta_s - \tau_i}{\sigma_i} \right)$. If it is then set that $\alpha_i = \frac{\lambda_i}{\sigma_i}$ and $b_i = -\frac{\tau_i}{\sigma_i}$, where $\sigma_i = \sqrt{1 - \Sigma \lambda_i^2}$, then the equation can be further rewritten to the MIRT two-parameter normal-ogive model $P(X_{is} = 1|\theta) = \Phi[\alpha_i \theta_s + b_i]$, in which $\alpha_i$ and $b_i$ are the discrimination and difficulty parameters, respectively. Thus, IRT discrimination and difficulty parameters can be transformed into CFA factor loading and threshold parameters, and vice versa (McLeod et al., 2001). The following equations can be used to derive MIRT parameters from CFA parameters and CFA parameters from MIRT parameters: $\lambda_i = \frac{\alpha_i}{\sqrt{1 + \Sigma \alpha_i^2}}$ and $\tau_i = -\frac{b_i}{\sqrt{1 + \Sigma \alpha_i^2}}$. In addition, when the item response categories are polytomous, as opposed to dichotomous, as is the case in the NSSE, the equation $b_i = \frac{\tau_i}{\lambda_i}$ can be used to transform the CFA multiple threshold parameters to the GRM model as: $b_{ij} = \frac{\tau_{ij}}{\lambda_i} (j = 1, \ldots, m_i$ number of score levels). This equivalency also relates to model identification; in both CFA and MIRT for a model to be estimated, the number of estimated parameters cannot be more than the amount of information in the variance/covariance matrix (Long, 1983; Svetina, 2011). This study utilized
both CFA and MIRT methodologies to investigate the psychometric properties of the revised NSSE and further explore whether the results of these separate methods were coherent and converged.

Despite the statistical equivalency of CFA and MIRT, they are distinct from each other in three ways: goal, focus, and methods (Reckase, 2009). First, the goal of factor analysis is defining factors, while the goal of MIRT is modeling the interactions between items and people. Second, in MIRT, there is a focus in the analysis on the individual characteristics of items, such as difficulty and discrimination, while factor analysis often does not focus on the individual characteristics of each item. Third, MIRT procedures work towards finding solutions that use a common latent space (i.e., common coordinate system) across instruments and samples so that “items will have parameter estimates on common metrics” which allows for instrument equating and linking; factor analysis does have some methods for putting solutions into a common latent space, however, those methods are not widely used (Reckase, 2009, p. 71).

**Summary**

Student engagement has been researched extensively with the college-aged population. Across the literature, there is considerable variation in the definition and measurement of student engagement. Sinatra et al.’s (2015) continuum of engagement measurement from person-oriented microlevel measures to context-oriented macrolevel measures provides a framework for researchers to think about how they are defining and measuring engagement. Student engagement is a complex, multidimensional construct, and as such, there are diverse methodologies that can be utilized when researching engagement. Each methodology and instrument has its own set of advantages and limitations. With self-report instruments, such as the NSSE, SERU, and NSS, a major advantage is their scalability (i.e., the ease with which this
type of instrument can be administered to large numbers of respondents). In order to address potential limitations, it is incumbent on researchers to conduct thorough psychometric analyses of these instruments. In addition, researchers and institutions should aim to utilize multiple methodologies to better capture the broad scope of the complex, multidimensional construct of student engagement.

There is a dearth of research investigating the psychometric properties and validity evidence of the revised NSSE. The proposed hierarchical structure between the ten engagement indicators and the four themes can be examined or confirmed using a higher-order factor analysis and MIRT. Through a review of the literature, no research was identified that investigated whether the proposed NSSE 2.0 four theme model is corroborated through higher-order factor analysis. In addition, no research was found that utilized IRT to analyze the NSSE 2.0, and only one study was found that examined the prior version of the NSSE utilizing IRT. This study addressed the research gap through the following research questions: (a) Based on unidimensional item response theory (IRT) and multidimensional item response theory (MIRT) criteria, are the NSSE items of the ten engagement indicators psychometrically sound?; (b) Based on confirmatory factor analysis (CFA) criteria, does the NSSE’s suggested factor structure of the construct of student engagement with ten engagement indicators organized into four themes fit the UHM 2015 NSSE data, and does a bifactor or second-order factor structure display a better fit?; and (c) What do the integrated results of (a) and (b) reveal about the psychometric adequacy of the NSSE 2.0?
Chapter 3: Methodology

Data Source

Pre-existing secondary data were obtained through the UHM’s Mānoa Institutional Research Office (MIRO). The data file with the 2015 UHM NSSE results was de-identified by MIRO staff prior to sharing the data with the researcher: all student names, birth dates, and other personally identifiable information were removed from the data. The MIRO staff inserted a unique identifier code for each respondent and the researcher does not have access to the file that allows for re-identification of respondents by cross-referencing to the unique identifier codes. The NSSE 2.0’s 47 engagement indicator items are presented in Table 2. The engagement indicators are further structured into higher level themes, which were presented in Table 1.

At the UHM, 5,677 freshmen and seniors were invited to participate in the NSSE in spring 2015. Of those invited students, 1,592 students completed the survey for a 32% response rate (Zhang, 2015). Of the respondents, 405 were freshman (sampling error +/-3.7%) and 1,187 were seniors (sampling error +/-2.2%). The sample was highly representative of the spring 2015 freshman and senior populations that were invited to take part in the survey (Zhang, 2015). When the survey was administered, the MIRO provided the list of UHM students to the NSSE staff at the Center for Postsecondary Research, Indiana University. The NSSE staff worked with the UHM to send out five survey invitation emails between February 10, 2015 and March 16, 2015; each email contained a unique survey link for each student. The survey site was closed to responses on June 1, 2015.
<table>
<thead>
<tr>
<th>Variable name</th>
<th>Item</th>
<th>Engagement Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Higher-Order Learning a</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOapply</td>
<td>Applying facts, theories, or methods to practical problems or new situations</td>
<td></td>
</tr>
<tr>
<td>HOanalyze</td>
<td>Analyzing an idea, experience, or line of reasoning in depth by examining its parts</td>
<td></td>
</tr>
<tr>
<td>HOevaluate</td>
<td>Evaluating a point of view, decision, or information source</td>
<td></td>
</tr>
<tr>
<td>HOform</td>
<td>Forming a new idea or understanding from various pieces of information</td>
<td></td>
</tr>
<tr>
<td><strong>Reflective and Integrative Learning b</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIntegrate</td>
<td>Combined ideas from different courses when completing assignments</td>
<td></td>
</tr>
<tr>
<td>RIsocietal</td>
<td>Connected your learning to societal problems or issues</td>
<td></td>
</tr>
<tr>
<td>RIdiverse</td>
<td>Included diverse perspectives (political, religious, racial/ethnic, gender, etc.) in course discussions or assignments</td>
<td></td>
</tr>
<tr>
<td>RIownview</td>
<td>Examined the strengths and weaknesses of your own views on a topic or issue</td>
<td></td>
</tr>
<tr>
<td>RIPerspect</td>
<td>Tried to better understand someone else's views by imagining how an issue looks from his or her perspective</td>
<td></td>
</tr>
<tr>
<td>RInewview</td>
<td>Learned something that changed the way you understand an issue or concept</td>
<td></td>
</tr>
<tr>
<td>RIconnect</td>
<td>Connected ideas from your courses to your prior experiences and knowledge</td>
<td></td>
</tr>
<tr>
<td><strong>Quantitative Reasoning b</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QRconclude</td>
<td>Reached conclusions based on your own analysis of numerical information (numbers, graphs, statistics, etc.)</td>
<td></td>
</tr>
<tr>
<td>QRproblem</td>
<td>Used numerical information to examine a real-world problem or issue (unemployment, climate change, public health, etc.)</td>
<td></td>
</tr>
<tr>
<td>QRevaluate</td>
<td>Evaluated what others have concluded from numerical information</td>
<td></td>
</tr>
<tr>
<td>Learning Strategies b</td>
<td>During the current school year, about how often have you done the following?</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>LSreading</td>
<td>Identified key information from reading assignments</td>
<td></td>
</tr>
<tr>
<td>LSnotes</td>
<td>Reviewed your notes after class</td>
<td></td>
</tr>
<tr>
<td>LSsummary</td>
<td>Summarized what you learned in class or from course materials</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collaborative Learning b</th>
<th>During the current school year, about how often have you done the following?</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLaskhelp</td>
<td>Asked another student to help you understand course material</td>
</tr>
<tr>
<td>CLexplain</td>
<td>Explained course material to one or more students</td>
</tr>
<tr>
<td>CLstudy</td>
<td>Prepared for exams by discussing or working through course material with other students</td>
</tr>
<tr>
<td>CLproject</td>
<td>Worked with other students on course projects or assignments</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discussions with Diverse Others b</th>
<th>During the current school year, about how often have you had discussions with people from the following groups?</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDrace</td>
<td>People of a race or ethnicity other than your own</td>
</tr>
<tr>
<td>DDeconomic</td>
<td>People from an economic background other than your own</td>
</tr>
<tr>
<td>DDreligion</td>
<td>People with religious beliefs other than your own</td>
</tr>
<tr>
<td>DDpolitical</td>
<td>People with political views other than your own</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student-Faculty Interaction b</th>
<th>During the current school year, about how often have you done the following?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFccareer</td>
<td>Talked about career plans with a faculty member</td>
</tr>
<tr>
<td>SOtherwork</td>
<td>Worked with a faculty member on activities other than coursework (committees, student groups, etc.)</td>
</tr>
<tr>
<td>SFdiscuss</td>
<td>Discussed course topics, ideas, or concepts with a faculty member outside of class</td>
</tr>
<tr>
<td>SFPperform</td>
<td>Discussed your academic performance with a faculty member</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effective Teaching Practices a</th>
<th>During the current school year, to what extent have your instructors done the following?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETgoals</td>
<td>Clearly explained course goals and requirements</td>
</tr>
<tr>
<td>ETorganize</td>
<td>Taught course sessions in an organized way</td>
</tr>
<tr>
<td>ETexample</td>
<td>Used examples or illustrations to explain difficult points</td>
</tr>
</tbody>
</table>
ETdraftfb  Provided feedback on a draft or work in progress
ETfeedback Provided prompt and detailed feedback on tests or completed assignments

Quality of Interactions

Indicate the quality of your interactions with the following people at your institution.

<table>
<thead>
<tr>
<th>Question</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIstudent</td>
<td>Students</td>
</tr>
<tr>
<td>QIadvisor</td>
<td>Academic advisors</td>
</tr>
<tr>
<td>QIfaculty</td>
<td>Faculty</td>
</tr>
<tr>
<td>QIstaff</td>
<td>Student services staff (career services, student activities, housing, etc.)</td>
</tr>
<tr>
<td>QIadmin</td>
<td>Other administrative staff and offices (registrar, financial aid, etc.)</td>
</tr>
</tbody>
</table>

Supportive Environment

How much does your institution emphasize the following?

<table>
<thead>
<tr>
<th>Question</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEacademic</td>
<td>Providing support to help students succeed academically</td>
</tr>
<tr>
<td>SElearsup</td>
<td>Using learning support services (tutoring services, writing center, etc.)</td>
</tr>
<tr>
<td>SEdiverse</td>
<td>Encouraging contact among students from different backgrounds (social, racial/ethnic, religious, etc.)</td>
</tr>
<tr>
<td>SESocial</td>
<td>Providing opportunities to be involved socially</td>
</tr>
<tr>
<td>SEwellness</td>
<td>Providing support for your overall well-being (recreation, health care, counseling, etc.)</td>
</tr>
<tr>
<td>SEnonacad</td>
<td>Helping you manage your non-academic responsibilities (work, family, etc.)</td>
</tr>
<tr>
<td>SEactivities</td>
<td>Attending campus activities and events (performing arts, athletic events, etc.)</td>
</tr>
<tr>
<td>SEevents</td>
<td>Attending events that address important social, economic, or political issues</td>
</tr>
</tbody>
</table>

\[a\] Response Categories: (1 = Very little, 2 = Some, 3 = Quite a bit, 4 = Very much)

\[b\] Response Categories: (1 = Never, 2 = Sometimes, 3 = Often, 4 = Very often)

\[c\] Response Categories: 1 (Poor) to 7 (Excellent)
Data Cleaning

Prior to data analysis, the data set was cleaned by applying the NSSE Engagement Indicator inclusion criteria. The inclusion criteria were as follows:

- Higher order learning (HO): answer all 4 items
- Reflective and interactive learning (RI): answer at least 6 of the 7 items
- Quantitative reasoning (QR): answer all 3 items
- Learning strategies (LS): answer all 3 items
- Collaborative learning (CL): answer all 4 items
- Discussions with diverse others (DD): answer all 4 items
- Student-faculty interaction (SF): answer all 4 items
- Effective teaching practice (ET): answer at least 4 of the 5 items
- Quality of interactions (QI): answer at least 4 of the 5 items
- Supportive environment (SE): answer at least 7 of the 8 items

Missing values were coded as -9 in the data set. Applying these criteria, the data sample was reduced to 1430 cases which consisted of 353 freshmen and 1077 seniors.

Data Analysis

The psychometric analysis had multiple phases (see Figure 6): (1) EFA analyses to explore a structure of the ten engagement indicators along with the pattern of factor loadings and identify which variables were most strongly correlated with each construct, (2) IRT analyses examining the psychometric soundness of the ten engagement indicator NSSE items utilizing unidimensional IRT and testing the fit of the suggested NSSE structure utilizing a MIRT model for each theme, and (3) CFA analyses to verify the fit of a ten engagement indicator structure, and four theme and ten engagement indicator bifactor and second-order factor structures. While
the study was originally designed to include the testing of a third-order structure, due to non-convergence of the second-order factor model for several themes, examining a third-order structure was not warranted. IRT analysis was used to scrutinize the specific item characteristics such as item and test information, and categorical response functions in each item. Exploratory and confirmatory factor analysis was used to test the proposed or theoretical structure in NSSE. The results of the IRT and factor analyses were then integrated to provide further clarity into the psychometric adequacy of the NSSE 2.0.
Figure 6

Data Analysis Flowchart

Data Cleaning

EFA – All items

EFA – 4 NSSE Themes

Unidimensional IRT - 10 Engagement Indicators

Multidimensional IRT – 4 NSSE Themes

Testing for Configural Invariance: Freshmen vs Seniors (by theme)

1st Order CFA – Items loadings onto Engagement Indicators (by Theme)

2nd Order CFA – Items loading on Engagement Indicators, then loading onto Theme

Bifactor CFA – Items loading onto both Engagement Indicator and Theme

Integration of IRT & CFA Results
These methods align with the theoretical framework of the study which conceptualizes student engagement as a multidimensional construct. This framework draws upon two key theories of student engagement: Fredricks et al.’s (2004) tripartite model of student engagement and Kuh et al.’s (2007b) framework of student success. The Fredricks et al. (2004) tripartite model classifies engagement into three dimensions: behavioral, cognitive, and emotional. The Kuh et al. (2007b) framework of student success incorporates antecedents and outcomes of student success with student engagement (including student and institutional factors) at the center.

These data analyses were performed using the software programs flexMIRT (Cai & Houts, 2017) and MPlus (Muthén & Muthén, 1998 - 2017). FlexMIRT is a robust IRT program capable of unidimensional and multidimensional item analysis on a variety of item-level models, and is capable of running confirmatory and exploratory factor analysis. MPlus is a flexible and robust statistical modeling program that offers a wide choice of models, estimators, and algorithms, and has an easy-to-use interface and graphical displays of data. In flexMIRT, Expectation Maximization (EM) and Metropolis-Hastings Robbins-Monro (MH-RM) algorithms were utilized for obtaining marginal maximum likelihood parameter estimates, with EM being the default algorithm and MH-RM being appropriate for higher-dimensional models with more than three latent variables (Jiang et al., 2016). The main difference between EM and MH-RM is their speed, not accuracy, in parameter estimation for higher-dimensional models (Cai, 2010). In MPlus, the weighted least square mean and variance adjusted (WLSMV) estimators were utilized for the confirmatory factor analyses. One limitation of flexMIRT is that it does not have graphic plotting capabilities; therefore, the IRTPRO (VPG, 2020a) software package was utilized to
produce graphic plots of the item level diagnostic statistics and total information curves for each engagement indicator.

**Model Evaluation**

In order to assess the goodness of fit of hypothesized models to actual data, multiple indices of fit should be evaluated to determine the acceptability of fit for each specified model (Brown, 2006; Crocker & Algina, 1986). In CFA, the fit of a model is considered “a global measure of the empirical relationships among all observed variables compared to the relationships implied by the structure of a theoretical model (model-implied covariance matrix) (Wang, 2012, p. 33).” The following model fit indices were utilized to examine the goodness of fit of the proposed models: $\chi^2$ statistic, $G^2$ statistic, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), likelihood ratio statistic, $-2\log$likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The $\chi^2$ statistic is an index of absolute fit (i.e., one that does not use an alternative model for comparison) with statistically significant values indicating that a hypothesized model has poor fit (i.e., does not fit the data well); a limitation of this fit statistic is that it is sensitive to sample size with larger sample sizes producing larger $\chi^2$ values leading to statistically significant results, which is not desirable for the purpose of goodness of fit test (Furr & Bacharach, 2014). The $G^2$ statistic is a loglikelihood ratio fit index which is less sensitive to sample size than the $\chi^2$ statistic. The RMSEA index adjusts for model complexity and indicates the extent to which a model fits reasonably well, with models that have fewer freely estimated parameters producing better fit; the RMSEA is not sensitive to sample size. The CFI and TLI are comparative fit indices which account for model complexity and that indicate the extent to which a hypothesized model fits in comparison to a null model with no relationships among variables. For the current study, RMSEA values less than or equal to .08 are considered acceptable, with RMSEA values of .06 or less being ideal; CFI and TLI values
greater than .90 are considered acceptable, with CFI and TLI values greater than .95 being ideal. The likelihood ratio statistic is a test of the sufficiency of a simpler model versus a more complex model. The -2loglikelihood statistic is a global fit index that indicates how well a model fits the data and represents the sum of the probabilities associated with the hypothesized model and actual data. The BIC and AIC fit indices are comparative fit indices that account for model complexity. The -2loglikelihood, AIC, and BIC indices can be used to compare the fit of models by examining the relative change in values across models, with the model having the lowest -2loglikelihood, AIC, and BIC values representing the best fitting model.

While IRT also evaluates overall model fit using $\chi^2$ statistic, RMSEA, -2loglikelihood, AIC, and BIC, there is an additional focus on fit at the item and person levels. The following item fit indices were examined for the proposed models within the IRT framework: marginal $\chi^2$ statistic for each item, local dependence (LD) for each item pair (LD $\chi^2$ values greater than 10 indicated a violation of the local independence assumption) and item-fit index $S – \chi^2$ to allow for the examination of fit at the item and person levels (VPG, 2020b). In flexMIRT, for items with ordered categorical responses, phi correlations are computed for both model implied and observed bivariate correlations and the LD $\chi^2$ values are followed by either the letter “p” or the letter “n.”

If the model implied correlation is lower than the observed correlation for a given item pair, a “p” is printed after the calculated $\chi^2$, meaning “positive” LD. If the model implied correlation is higher, an “n” is printed, indicating “negative” LD. (Cai & Houts, 2017, p. 18)

Violations of local independence would indicate that the items within a scale did not meet the IRT modeling assumption that the only influence on an individual’s item response is that of the
latent trait variable being measured and that item responses are independent from other item responses (Toland, 2014).

**Exploratory Factor Analysis**

The preliminary analysis conducted was exploratory factor analysis in order to explore the pattern of factor loadings and identify which variables were correlated with each construct. Due to the ordered categorical nature of the NSSE item scales, the EFA analyses were conducted by specifying which variables were categorical in the MPlus syntax with the WLSMV estimator. This initial EFA was conducted using the MPlus software program, with the WLSMV estimator, by running separate EFA analyses for the items comprising each of the four NSSE themes and running the EFA analysis on all 47 items together. To facilitate interpretation of the factors, an oblique geomin rotation was utilized in MPlus. Factors loading values greater than 0.30 were identified for each construct.

**Item Response Theory**

Unidimensional IRT analyses were conducted on each of the ten engagement indicators using Samejima’s (1969) GRM model for the ordered categorical responses of the NSSE items, using the flexMIRT software package. Goodness of fit, local dependence, and item parameter estimates, including threshold parameters and item discrimination, were examined. In addition, the IRTPRO software package was utilized for item level diagnostic statistics, including item category curves and information functions for each item, as well as total information curves for each engagement indicator.

Multidimensional IRT analyses were also conducted with the flexMIRT software package for each of the four NSSE themes, with the items loading onto their engagement
indicators, using Samejima’s (1969) GRM model. Goodness of fit, local dependence, and item parameter estimates, including threshold parameters and item discrimination, were examined.

**Confirmatory Factor Analysis**

The following CFA analyses were conducted using the MPlus program.

**Testing for Configural Invariance.** To test configural invariance across the two participant groups (freshmen and seniors) of the NSSE, configural invariance analysis was conducted with the MPlus software program to determine if further analyses could be conducted on a merged sample with both groups. Configural invariance analysis is an extension of confirmatory factor analysis in which the data is set into groups, the model fit for each group is calculated separately, and then a multi-group model fit is calculated. The results of a configural invariance analysis allow the researcher to determine if the different groups interpret the measure in similar ways.

**First Order Confirmatory Factor Analysis.** The first CFA models analyzed were first order models for each of the four NSSE themes, with the items loading onto their engagement indicator that correspond to each theme, but not loading onto a theme (see Figures 7 – 10). These models sought to test the number of factors and relationship patterns between factors and items based on the a priori hypothesis that the NSSE items measure constructs grouped into engagement indicators and themes. The CFA models were specified to reflect the hypothesized relationship between items, factors that represent the latent traits the items are supposed to be measuring, and measurement error that accounts for all the unique sources of variance that are not explained by the factors (Long, 1983).
Figure 7

Academic Challenge CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR)
Figure 8

Learning with Peers CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD)
Figure 9

Experiences with Faculty CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET)
Second-Order Confirmatory Factor Analysis. The second order CFA models were analyzed for each of the four NSSE themes, with the hierarchical structure of items loading onto their engagement indicator, and the engagement indicators loading onto their theme (see Figures 11 – 14).
Figure 11

Academic Challenge (AC) 2nd Order CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR)
Figure 12

Learning with Peers (LP) 2nd Order CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD)
Figure 13

Experiences with Faculty (EF) 2nd Order CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET)
Bifactor Confirmatory Factor Analysis. The bifactor CFA models were analyzed for each of the four NSSE themes with each item simultaneously loading onto one general latent dimension (the theme), as well as loading onto a specific (secondary) latent dimension (the engagement indicators) (see Figures 15 – 18).
Academic Challenge (AC) Bifactor CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR)
Figure 16

*Learning with Peers (LP) Bifactor CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD)*
Figure 17

Experiences with Faculty (EF) Bifactor CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET)
Model Comparison. After running the 1st order, 2nd order, and bifactor CFA models, the $\chi^2$ difference test was utilized to compare two nested models by calculating the adjusted difference between their $\chi^2$ statistics in Mplus. The $\chi^2$ difference test analysis was conducted to compare the Academic Challenge 1st order model versus 2nd order model. The Academic Challenge bifactor model could not be compared versus the 2nd order model because they are not
nested models; these were compared by examining goodness of fit indices. In addition, the Learning with Peers, Experiences with Faculty, and Campus Environment models could not be compared using the $\chi^2$ difference tests analyses due to the nonconvergence of various models and non-nested models. The models which could not be compared using the $\chi^2$ difference test were compared by examining goodness of fit indices.
Chapter 4: Results

Exploratory Factor Analysis

The EFA analyses were done on all 47 NSSE items together and also by the four NSSE themes.

All 47 Items

The EFA analysis, using the WLSMV estimator and oblique geomin rotation, was utilized on all 47 NSSE items together resulted in a ten-factor solution, which corresponded with the ten engagement indicator groupings proposed by the NSSE, as shown in Table 3: Higher-Order Learning (λ = 0.74 – 0.89), Reflective and Integrative Learning (λ = 0.59 – 0.81), Quantitative Reasoning (λ = 0.78 – 0.89), Learning Strategies (λ = 0.41 – 0.86), Collaborative Learning (λ = 0.61 – 0.83), Discussions with Diverse Others (λ = 0.86 – 0.92), Student-Faculty Interactions (λ = 0.68 – 0.80), Effective Teaching Practices (λ = 0.69 – 0.85), Quality of Interactions (λ = 0.38 – 0.86), and Supportive Environment (λ = 0.59 – 0.78). The items with the lowest loadings were LSreading (λ = 0.41) and QIstudent (λ = 0.38).
### Table 3

**All 47 NSSE Items Factor Loadings of EFA**

<table>
<thead>
<tr>
<th>Item</th>
<th>(\lambda_1)</th>
<th>(\lambda_2)</th>
<th>(\lambda_3)</th>
<th>(\lambda_4)</th>
<th>(\lambda_5)</th>
<th>(\lambda_6)</th>
<th>(\lambda_7)</th>
<th>(\lambda_8)</th>
<th>(\lambda_9)</th>
<th>(\lambda_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOapply</td>
<td>0.774*</td>
<td>-0.055</td>
<td>0.091*</td>
<td>-0.007</td>
<td>0.065*</td>
<td>0.016</td>
<td>0.015</td>
<td>-0.020</td>
<td>-0.017</td>
<td>0.034</td>
</tr>
<tr>
<td>HOanalyze</td>
<td>0.890*</td>
<td>-0.012</td>
<td>0.043*</td>
<td>0.004</td>
<td>0.028</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.025</td>
<td>0.005</td>
<td>0.045*</td>
</tr>
<tr>
<td>HOevaluate</td>
<td>0.789*</td>
<td>0.123*</td>
<td>-0.054*</td>
<td>0.007</td>
<td>-0.047*</td>
<td>-0.020</td>
<td>0.013</td>
<td>0.064*</td>
<td>0.060*</td>
<td>-0.030</td>
</tr>
<tr>
<td>HOform</td>
<td>0.740*</td>
<td>0.138*</td>
<td>-0.006</td>
<td>0.009</td>
<td>-0.036*</td>
<td>0.009</td>
<td>0.035</td>
<td>0.069*</td>
<td>0.032</td>
<td>-0.047*</td>
</tr>
<tr>
<td>RIntegrate</td>
<td>0.049</td>
<td>0.594*</td>
<td>0.061*</td>
<td>0.017</td>
<td>0.098*</td>
<td>0.010</td>
<td>0.032</td>
<td>-0.028</td>
<td>0.071*</td>
<td>0.001</td>
</tr>
<tr>
<td>RIsocietal</td>
<td>0.004</td>
<td>0.784*</td>
<td>0.041*</td>
<td>0.017</td>
<td>-0.093*</td>
<td>0.028</td>
<td>0.023</td>
<td>-0.051*</td>
<td>0.095*</td>
<td>0.012</td>
</tr>
<tr>
<td>RIdiverse</td>
<td>-0.058*</td>
<td>0.791*</td>
<td>-0.062*</td>
<td>0.054*</td>
<td>-0.143*</td>
<td>-0.014</td>
<td>0.042*</td>
<td>0.024</td>
<td>0.125*</td>
<td>-0.003</td>
</tr>
<tr>
<td>RIownview</td>
<td>0.024</td>
<td>0.804*</td>
<td>0.038</td>
<td>-0.021</td>
<td>0.021</td>
<td>-0.054*</td>
<td>-0.033</td>
<td>0.020</td>
<td>0.053*</td>
<td>0.012</td>
</tr>
<tr>
<td>RIperspect</td>
<td>-0.026</td>
<td>0.809*</td>
<td>0.044</td>
<td>-0.031</td>
<td>0.084*</td>
<td>-0.017</td>
<td>-0.044</td>
<td>0.032</td>
<td>-0.070*</td>
<td>0.028</td>
</tr>
<tr>
<td>RIconnect</td>
<td>0.126*</td>
<td>0.730*</td>
<td>-0.017</td>
<td>-0.039</td>
<td>0.071*</td>
<td>0.045</td>
<td>0.001</td>
<td>0.031</td>
<td>-0.057*</td>
<td>0.001</td>
</tr>
<tr>
<td>RIconnect</td>
<td>0.081*</td>
<td>0.711*</td>
<td>0.003</td>
<td>0.019</td>
<td>0.050*</td>
<td>0.103*</td>
<td>0.031</td>
<td>0.017</td>
<td>-0.052*</td>
<td>-0.014</td>
</tr>
<tr>
<td>QRconclude</td>
<td>0.045*</td>
<td>-0.038*</td>
<td>0.783*</td>
<td>0.006</td>
<td>0.087*</td>
<td>0.006</td>
<td>0.031</td>
<td>0.000</td>
<td>-0.046*</td>
<td>-0.001</td>
</tr>
<tr>
<td>QRproblem</td>
<td>-0.017</td>
<td>0.047*</td>
<td>0.887*</td>
<td>0.006</td>
<td>-0.081*</td>
<td>0.012</td>
<td>0.006</td>
<td>0.011</td>
<td>0.057*</td>
<td>0.001</td>
</tr>
<tr>
<td>QRevaluate</td>
<td>0.011</td>
<td>0.048*</td>
<td>0.837*</td>
<td>0.021</td>
<td>0.020*</td>
<td>0.013</td>
<td>-0.006</td>
<td>0.016</td>
<td>0.029*</td>
<td>-0.012</td>
</tr>
<tr>
<td>LSreading</td>
<td>0.120*</td>
<td>0.171*</td>
<td>0.049</td>
<td>0.189*</td>
<td>-0.047*</td>
<td>0.161*</td>
<td>0.405*</td>
<td>-0.061*</td>
<td>-0.036</td>
<td>0.026</td>
</tr>
<tr>
<td>LSnotes</td>
<td>0.018</td>
<td>-0.053*</td>
<td>0.005</td>
<td>0.015</td>
<td>0.068*</td>
<td>0.023</td>
<td>0.793*</td>
<td>0.032*</td>
<td>0.017</td>
<td>0.016</td>
</tr>
<tr>
<td>LSsummary</td>
<td>0.007</td>
<td>0.058*</td>
<td>-0.003</td>
<td>-0.010</td>
<td>0.033*</td>
<td>0.003</td>
<td>0.858*</td>
<td>0.027</td>
<td>0.003</td>
<td>0.005</td>
</tr>
<tr>
<td>CLaskhelp</td>
<td>-0.040</td>
<td>-0.015</td>
<td>-0.026</td>
<td>-0.026</td>
<td>0.769*</td>
<td>-0.052*</td>
<td>0.064*</td>
<td>0.036</td>
<td>0.010</td>
<td>0.006</td>
</tr>
<tr>
<td>CLexplain</td>
<td>-0.012</td>
<td>0.185*</td>
<td>0.050*</td>
<td>0.035</td>
<td>0.622*</td>
<td>-0.017</td>
<td>0.113*</td>
<td>-0.157*</td>
<td>0.053*</td>
<td>-0.013</td>
</tr>
<tr>
<td>CLstudy</td>
<td>0.014</td>
<td>0.010</td>
<td>-0.001</td>
<td>0.003</td>
<td>0.833*</td>
<td>0.002</td>
<td>0.022</td>
<td>-0.002</td>
<td>0.081*</td>
<td>-0.015</td>
</tr>
<tr>
<td>CLproject</td>
<td>0.087*</td>
<td>0.071*</td>
<td>0.049</td>
<td>0.052*</td>
<td>0.609*</td>
<td>0.007</td>
<td>-0.076*</td>
<td>0.031</td>
<td>0.041</td>
<td>-0.024</td>
</tr>
<tr>
<td>DDrace</td>
<td>-0.010</td>
<td>0.063*</td>
<td>0.006</td>
<td>0.856*</td>
<td>0.027</td>
<td>0.044*</td>
<td>-0.005</td>
<td>-0.043*</td>
<td>-0.084*</td>
<td>0.073*</td>
</tr>
<tr>
<td>DEeconomic</td>
<td>0.014</td>
<td>0.000</td>
<td>0.053*</td>
<td>0.882*</td>
<td>0.012</td>
<td>0.009</td>
<td>0.024</td>
<td>-0.016</td>
<td>-0.034*</td>
<td>0.020</td>
</tr>
<tr>
<td>DDrigion</td>
<td>-0.007</td>
<td>-0.025</td>
<td>-0.028</td>
<td>0.915*</td>
<td>0.015</td>
<td>-0.019</td>
<td>-0.001</td>
<td>0.072*</td>
<td>0.054*</td>
<td>-0.034*</td>
</tr>
<tr>
<td>DDpolitical</td>
<td>0.002</td>
<td>-0.029</td>
<td>0.010</td>
<td>0.907*</td>
<td>-0.015</td>
<td>-0.051*</td>
<td>0.010</td>
<td>0.051*</td>
<td>0.069*</td>
<td>-0.024</td>
</tr>
<tr>
<td>SFcareer</td>
<td>0.062*</td>
<td>0.010</td>
<td>0.002</td>
<td>-0.001</td>
<td>0.072*</td>
<td>0.044</td>
<td>-0.007</td>
<td>-0.010</td>
<td>0.723*</td>
<td>0.047*</td>
</tr>
<tr>
<td>SFotherwork</td>
<td>0.002</td>
<td>-0.004</td>
<td>0.097*</td>
<td>0.001</td>
<td>0.013</td>
<td>-0.036</td>
<td>-0.064*</td>
<td>0.070*</td>
<td>0.771*</td>
<td>0.010</td>
</tr>
<tr>
<td>SFdiscuss</td>
<td>0.028</td>
<td>0.022</td>
<td>0.044*</td>
<td>0.018</td>
<td>0.054*</td>
<td>0.010</td>
<td>0.021</td>
<td>-0.021</td>
<td>0.795*</td>
<td>0.004</td>
</tr>
<tr>
<td>SFperform</td>
<td>0.023</td>
<td>0.063*</td>
<td>0.006</td>
<td>-0.015</td>
<td>0.143*</td>
<td>0.042</td>
<td>0.034</td>
<td>0.016</td>
<td>0.675*</td>
<td>0.040</td>
</tr>
<tr>
<td>ETgoals</td>
<td>0.025</td>
<td>0.052*</td>
<td>0.006</td>
<td>0.006</td>
<td>0.010</td>
<td>0.744*</td>
<td>0.007</td>
<td>-0.004</td>
<td>-0.005</td>
<td>0.073*</td>
</tr>
<tr>
<td>EToorganize</td>
<td>0.002</td>
<td>0.006</td>
<td>0.033</td>
<td>0.008</td>
<td>-0.028</td>
<td>0.849*</td>
<td>-0.031</td>
<td>-0.004</td>
<td>-0.049*</td>
<td>0.028</td>
</tr>
<tr>
<td>ETexample</td>
<td>0.036</td>
<td>0.014</td>
<td>0.018</td>
<td>0.036</td>
<td>0.050*</td>
<td>0.776*</td>
<td>0.006</td>
<td>0.018</td>
<td>0.017</td>
<td>0.019</td>
</tr>
<tr>
<td>ETDraftfb</td>
<td>-0.017</td>
<td>0.011</td>
<td>-0.011</td>
<td>-0.027</td>
<td>-0.029</td>
<td>0.690*</td>
<td>0.061*</td>
<td>0.109*</td>
<td>0.219*</td>
<td>-0.059*</td>
</tr>
<tr>
<td>ETfeedback</td>
<td>-0.049*</td>
<td>-0.037</td>
<td>0.016</td>
<td>-0.027</td>
<td>-0.001</td>
<td>0.796*</td>
<td>0.064*</td>
<td>0.054*</td>
<td>0.168*</td>
<td>-0.021</td>
</tr>
<tr>
<td>QIstudent</td>
<td>-0.008</td>
<td>0.002</td>
<td>-0.031</td>
<td>0.126*</td>
<td>0.269*</td>
<td>0.095*</td>
<td>-0.021</td>
<td>0.102*</td>
<td>-0.055</td>
<td>0.380*</td>
</tr>
<tr>
<td>QIadvisor</td>
<td>0.008</td>
<td>-0.046</td>
<td>-0.014</td>
<td>0.001</td>
<td>0.026</td>
<td>0.126*</td>
<td>-0.068*</td>
<td>-0.019</td>
<td>0.186*</td>
<td>0.647*</td>
</tr>
<tr>
<td>QIfaculty</td>
<td>0.044*</td>
<td>0.056*</td>
<td>-0.061*</td>
<td>0.089*</td>
<td>-0.057*</td>
<td>0.253*</td>
<td>0.002</td>
<td>-0.079*</td>
<td>0.217*</td>
<td>0.577*</td>
</tr>
<tr>
<td>QIstaff</td>
<td>-0.041</td>
<td>0.015</td>
<td>0.038</td>
<td>-0.044*</td>
<td>0.017</td>
<td>-0.100*</td>
<td>0.046</td>
<td>0.103*</td>
<td>0.014</td>
<td>0.834*</td>
</tr>
<tr>
<td>QIadmin</td>
<td>-0.004</td>
<td>-0.002</td>
<td>0.026</td>
<td>-0.010</td>
<td>-0.060*</td>
<td>-0.039*</td>
<td>0.081*</td>
<td>0.049*</td>
<td>-0.011</td>
<td>0.860*</td>
</tr>
<tr>
<td>SEacademic</td>
<td>0.108*</td>
<td>-0.022</td>
<td>0.016</td>
<td>-0.023</td>
<td>0.103*</td>
<td>0.185*</td>
<td>-0.047*</td>
<td>0.587*</td>
<td>-0.088*</td>
<td>0.155*</td>
</tr>
<tr>
<td>SElearnSup</td>
<td>0.084*</td>
<td>-0.045</td>
<td>0.030</td>
<td>-0.021</td>
<td>0.089*</td>
<td>0.116*</td>
<td>-0.052*</td>
<td>0.637*</td>
<td>-0.129*</td>
<td>0.081*</td>
</tr>
<tr>
<td>SEdiverse</td>
<td>0.083*</td>
<td>0.009</td>
<td>-0.019</td>
<td>0.051*</td>
<td>-0.012</td>
<td>0.020</td>
<td>0.037</td>
<td>0.681*</td>
<td>0.041</td>
<td>-0.003</td>
</tr>
<tr>
<td>SEsocial</td>
<td>-0.010</td>
<td>0.087*</td>
<td>-0.029</td>
<td>0.068*</td>
<td>0.057*</td>
<td>0.060*</td>
<td>-0.054*</td>
<td>0.761*</td>
<td>0.024</td>
<td>-0.016</td>
</tr>
<tr>
<td>SEwellness</td>
<td>-0.028</td>
<td>0.095*</td>
<td>-0.023</td>
<td>0.020</td>
<td>-0.003</td>
<td>0.033</td>
<td>0.028</td>
<td>0.778*</td>
<td>-0.017</td>
<td>0.015</td>
</tr>
<tr>
<td>SENonacad</td>
<td>-0.043</td>
<td>-0.068*</td>
<td>0.065*</td>
<td>-0.067*</td>
<td>-0.027</td>
<td>-0.027</td>
<td>0.090*</td>
<td>0.763*</td>
<td>0.131*</td>
<td>0.009</td>
</tr>
<tr>
<td>SEactivities</td>
<td>-0.022</td>
<td>0.069*</td>
<td>-0.011</td>
<td>0.049*</td>
<td>-0.032</td>
<td>-0.026</td>
<td>0.022</td>
<td>0.729*</td>
<td>0.000</td>
<td>-0.011</td>
</tr>
<tr>
<td>SEevents</td>
<td>0.036</td>
<td>0.025</td>
<td>0.034</td>
<td>0.021</td>
<td>-0.121*</td>
<td>-0.047*</td>
<td>0.048*</td>
<td>0.768*</td>
<td>0.057*</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*Note. The extraction method was weighted least squares and variance (WLSMV) with an oblique geomin rotation. Factor loadings above .30 are in bold.

*p < .05
**Academic Challenge Theme**

The results of the EFA analyses for the seventeen Academic Challenge theme items corresponded with the engagement indicator groupings proposed by the NSSE, as shown in Table 4. Four factors were extracted, using the WLSMV estimator and oblique geomin rotation, that corresponded to the NSSE Engagement Indicators: Higher-Order Learning ($\lambda = 0.76 - 0.95$), Reflective and Integrative Learning ($\lambda = 0.66 - 0.86$), Quantitative Reasoning ($\lambda = 0.83 - 0.91$), and Learning Strategies ($\lambda = 0.48 - 0.92$). The item with the lowest loading was LSreading ($\lambda = 0.48$).

**Table 4**

*Academic Challenge Theme Factor Loadings of EFA*

<table>
<thead>
<tr>
<th>Item</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\lambda_3$</th>
<th>$\lambda_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOapply</td>
<td>0.02</td>
<td>0.04</td>
<td><strong>0.85</strong></td>
<td>-0.08</td>
</tr>
<tr>
<td>HOanalyze</td>
<td>-0.02</td>
<td>0.03</td>
<td><strong>0.95</strong></td>
<td>-0.04</td>
</tr>
<tr>
<td>HOevaluate</td>
<td>0.01</td>
<td>-0.05</td>
<td><strong>0.82</strong></td>
<td>0.10</td>
</tr>
<tr>
<td>HOform</td>
<td>0.03</td>
<td>0.01</td>
<td><strong>0.76</strong></td>
<td>0.12</td>
</tr>
<tr>
<td>RIntegrate</td>
<td>0.03</td>
<td>0.09</td>
<td>0.05</td>
<td><strong>0.66</strong></td>
</tr>
<tr>
<td>RLsocietal</td>
<td>0.01</td>
<td>0.04</td>
<td>-0.01</td>
<td><strong>0.83</strong></td>
</tr>
<tr>
<td>RIdiverse</td>
<td>0.04</td>
<td>-0.08</td>
<td>-0.07</td>
<td><strong>0.86</strong></td>
</tr>
<tr>
<td>RIfconn</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.00</td>
<td><strong>0.83</strong></td>
</tr>
<tr>
<td>RIperspect</td>
<td>-0.03</td>
<td>0.05</td>
<td>-0.04</td>
<td><strong>0.82</strong></td>
</tr>
<tr>
<td>RLnewview</td>
<td>-0.00</td>
<td>-0.02</td>
<td>0.13</td>
<td><strong>0.76</strong></td>
</tr>
<tr>
<td>RLconnect</td>
<td>0.04</td>
<td>-0.01</td>
<td>0.10</td>
<td><strong>0.75</strong></td>
</tr>
<tr>
<td>QRconclude</td>
<td>0.03</td>
<td><strong>0.83</strong></td>
<td>0.04</td>
<td>-0.07</td>
</tr>
<tr>
<td>QRproblem</td>
<td>0.00</td>
<td><strong>0.91</strong></td>
<td>-0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>QRevaluate</td>
<td>-0.01</td>
<td><strong>0.90</strong></td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>LSreading</td>
<td><strong>0.48</strong></td>
<td>0.03</td>
<td>0.13</td>
<td>0.20</td>
</tr>
<tr>
<td>LSnotes</td>
<td><strong>0.85</strong></td>
<td>0.01</td>
<td>0.01</td>
<td>-0.06</td>
</tr>
<tr>
<td>LSsummary</td>
<td><strong>0.92</strong></td>
<td>-0.00</td>
<td>-0.03</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*Note.* The extraction method was WLSMV with an oblique geomin rotation. Factor loadings above .30 are in bold.
**Learning with Peers Theme**

The results of the EFA analyses for the eight Learning with Peers theme items corresponded with the engagement indicator groupings proposed by the NSSE, as shown in Table 5. Two factors were extracted, using the WLSMV estimator and oblique geomin rotation, that corresponded to the NSSE Engagement Indicators: Collaborative Learning ($\lambda = -0.71$– 0.92) and Discussions with Diverse Others ($\lambda = 0.89$ – 0.92).

**Table 5**

*Learning with Peers Theme Factor Loadings of EFA*

<table>
<thead>
<tr>
<th>Item</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLaskhelp</td>
<td>-0.76</td>
<td>-0.04</td>
</tr>
<tr>
<td>CLexplain</td>
<td>-0.71</td>
<td>0.06</td>
</tr>
<tr>
<td>CLstudy</td>
<td>-0.92</td>
<td>-0.03</td>
</tr>
<tr>
<td>CLproject</td>
<td>-0.71</td>
<td>0.04</td>
</tr>
<tr>
<td>DDrace</td>
<td>0.01</td>
<td>0.89</td>
</tr>
<tr>
<td>DDeconomic</td>
<td>0.00</td>
<td>0.92</td>
</tr>
<tr>
<td>DDreligion</td>
<td>-0.01</td>
<td>0.92</td>
</tr>
<tr>
<td>DDPolitical</td>
<td>-0.00</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*Note.* The extraction method was WLSMV with an oblique geomin rotation. Factor loadings above .30 are in bold.

**Experiences with Faculty Theme**

The results of the EFA analyses for the nine Experiences with Faculty theme items corresponded with the engagement indicator groupings proposed by the NSSE, as shown in Table 6. Two factors were extracted, using the WLSMV estimator and oblique geomin rotation, that corresponded to the NSSE Engagement Indicators: Student-Faculty Interactions ($\lambda = -0.80$ – 0.88) and Effective Teaching Practices ($\lambda = 0.72$ – 0.89).
### Table 6

*Experiences with Faculty Theme Factor Loadings of EFA*

<table>
<thead>
<tr>
<th>Item</th>
<th>λ1</th>
<th>λ2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFcareer</td>
<td>0.04</td>
<td><strong>-0.80</strong></td>
</tr>
<tr>
<td>SFotherwork</td>
<td>-0.04</td>
<td><strong>-0.82</strong></td>
</tr>
<tr>
<td>SFDiscuss</td>
<td>-0.02</td>
<td><strong>-0.88</strong></td>
</tr>
<tr>
<td>SFperform</td>
<td>0.04</td>
<td><strong>-0.81</strong></td>
</tr>
<tr>
<td>ETgoals</td>
<td><strong>0.83</strong></td>
<td>0.01</td>
</tr>
<tr>
<td>ETorganize</td>
<td><strong>0.89</strong></td>
<td>0.11</td>
</tr>
<tr>
<td>ETexample</td>
<td><strong>0.86</strong></td>
<td>-0.00</td>
</tr>
<tr>
<td>ETdraftfb</td>
<td><strong>0.72</strong></td>
<td>-0.13</td>
</tr>
<tr>
<td>ETfeedback</td>
<td><strong>0.80</strong></td>
<td>-0.08</td>
</tr>
</tbody>
</table>

*Note.* The extraction method was WLSMV with an oblique geomin rotation. Factor loadings above .30 are in bold.

### Campus Environment Theme

The results of the EFA analysis for the thirteen Campus Environment theme items corresponded with the engagement indicator groupings proposed by the NSSE, as shown in Table 7. Two factors were extracted, using the WLSMV estimator and oblique geomin rotation, that corresponded to the NSSE Engagement Indicators: Quality of Interactions ($\lambda = -0.39 – 0.90$) and Supportive Environment ($\lambda = 0.69 – 0.86$). The item with the lowest loading was QIstudent ($\lambda = 0.39$).
Table 7

*Campus Environment Theme Factor Loadings of EFA*

<table>
<thead>
<tr>
<th>Item</th>
<th>λ1</th>
<th>λ2</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIstudent</td>
<td>0.20</td>
<td>-0.39</td>
</tr>
<tr>
<td>QIadvisor</td>
<td>0.00</td>
<td>-0.73</td>
</tr>
<tr>
<td>QIfaculty</td>
<td>0.04</td>
<td>-0.72</td>
</tr>
<tr>
<td>QIstaff</td>
<td>-0.00</td>
<td>-0.86</td>
</tr>
<tr>
<td>QIadmin</td>
<td>-0.03</td>
<td>-0.90</td>
</tr>
<tr>
<td>SEacademic</td>
<td>0.69</td>
<td>-0.14</td>
</tr>
<tr>
<td>SElearnsup</td>
<td>0.70</td>
<td>-0.05</td>
</tr>
<tr>
<td>SEdiverse</td>
<td>0.78</td>
<td>0.01</td>
</tr>
<tr>
<td>SEsocial</td>
<td>0.86</td>
<td>0.04</td>
</tr>
<tr>
<td>SEwellness</td>
<td>0.85</td>
<td>0.01</td>
</tr>
<tr>
<td>SENonacad</td>
<td>0.77</td>
<td>-0.00</td>
</tr>
<tr>
<td>SEactivities</td>
<td>0.77</td>
<td>0.04</td>
</tr>
<tr>
<td>SEevents</td>
<td>0.79</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Note.* The extraction method was WLSMV with an oblique geomin rotation. Factor loadings above .30 are in bold.

**Item Response Theory Analysis**

Unidimensional IRT analyses were conducted on each of the ten engagement indicators and multidimensional IRT analyses were conducted on each of the four NSSE themes.

**Academic Challenge Theme**

**Unidimensional IRT of Higher Order Learning Engagement Indicator.** The goodness of fit statistics of the unidimensional IRT analysis for the Higher Order Learning engagement indicator did not indicate goodness of fit based on the $\chi^2(239) = 9977.41, p < .01$ and the RMSEA value larger than .08; while the $G^2(111) = 791.41, p < .01$ and the RMSEA value of 0.07 test of model fit did fall within an acceptable value. Both the Pearson $\chi^2$ and likelihood ratio statistic $G^2$ tests of model fit were statistically significant which could indicate poor fit; note that this result also may be attributed to the large sample size. Item parameter
estimates indicated that the most discriminating item in the Higher Order Learning EI was HOanalyze ($a = 4.67$), while the least discriminating item was HOapply ($a = 2.5$). The point at which the descending curve reaches 0.50 probability is called the threshold 1 ($b_1$), and the point at which the next curve meets 0.50 is called the threshold 2 ($b_2$), et cetera. As shown in Figure 19, there were three threshold parameter estimates with four response categories, (1) Very little, (2) Some, (3) Quite a bit, and (4) Very much, for the Higher Order Learning EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the HOanalyze item had the highest threshold parameter value at the $b_1$ level ($b_1 = -1.75$), indicating that of the Higher Order Learning EI items, the HOanalyze item had a lower likelihood to be endorsed at the higher response value, “Some.” At the $b_3$ level, HOevaluate ($b_3 = 0.62$) had a lower likelihood to be endorsed at the higher response value, “Very much.” The standardized local dependence statistics indicated that all items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values > 10, ranging from 18.3 to 38.7. Most of the local dependence values for the Higher Order Learning EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

The category response curves and information functions for the Higher Order Learning engagement indicator items are shown in Figure 19 below. For all four items, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta
value where there was a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative item (indicated by the dashed information function line) was HOanalyze. The total information curve for the Higher Order Learning engagement indicator showed that the HO items can best be used from theta values of approximately -2.5 to 1 (See Figure 20). This indicated that the Higher Order Learning engagement indicator scores are least discriminating and informative for theta values above one (i.e., students who reported the most emphasis in their coursework on higher order learning practices).

**Figure 19**

*Category Response Curves and Information Functions for Higher Order Learning EI*

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.
Unidimensional IRT of Reflective and Integrative Learning Engagement Indicator.

The unidimensional IRT analysis goodness of fit Pearson $\chi^2$ statistics for the Reflective and Integrative Learning engagement indicator could not be computed; the underlying contingency table was too sparse to compute the full-information fit values because the total number of cross-classifications was $4^7 = 16384$, which was considerably larger than the sample size of 1430. The $G^2(619) = 3289.16, p < .01$ and the RMSEA value of 0.05 indicated good model fit; however, the $G^2$ values should be interpreted with caution given the issue with the sparse contingency table. Item parameter estimates indicated that the most discriminating item in the Reflective and Integrative Learning EI was RIsocietal ($a = 2.66$), while the least discriminating item was RIntegrate ($a = 1.88$). As shown in Figure 21, there were three threshold parameters with four response categories, (1) Never, (2) Sometimes, (3) Often, and (4) Very often, for the Reflective and Integrative Learning EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from...
lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the RIdiverse item had the highest threshold parameter value at all \( b \) levels (\( b_1 = -1.61, b_2 = -0.20, b_3 = 0.91 \)), indicating that of the Reflective and Integrative Learning EI items, the RIdiverse item had a lower likelihood to be endorsed at the higher response value for each response category threshold. The standardized local dependence statistics indicated that the majority of items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values > 10, ranging from 5.8 to 35.4. Approximately two-thirds of the local dependence values for the Reflective and Integrative Learning EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

The category response curves and information functions for the Reflective and Integrative Learning engagement indicator items are shown in Figure 21 below. For all seven items, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta value where there was a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative items (indicated by the dashed information function line) were RIsocietal and RIownview. The total information curve for the Reflective and Integrative Learning engagement indicator showed that the RI EI items can best be used from theta values of approximately -3 to 1.5 (See Figure 22). This indicated that the Reflective and Integrative Learning engagement indicator scores are least discriminating and informative for theta values above 1.5 (i.e., students
who reported the most frequent amounts of engagement in reflective and integrative learning practices).

**Figure 21**

*Category Response Curves and Information Functions for Reflective and Integrative Learning EI*

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.

**Figure 22**

*Total Information Curve for Reflective and Integrative Learning EI*
Unidimensional IRT of Quantitative Reasoning Engagement Indicator. The goodness of fit statistics of the unidimensional IRT analysis for the Quantitative Reasoning engagement indicator did not indicate goodness of fit based on the $\chi^2(51) = 1353.16, p < .01$ and the RMSEA value greater than .08; while the $G^2(43) = 354.54, p < .01$ and the RMSEA value of 0.07 test of model fit did fall within an acceptable value. Both the Pearson $\chi^2$ and likelihood ratio statistic $G^2$ tests of model fit were statistically significant which could indicate poor fit; note that this result also may be attributed to the large sample size. Item parameter estimates indicated that the most discriminating item in the Quantitative Reasoning EI was QRproblem ($a = 3.83$), while the least discriminating item was QRconclude ($a = 2.52$). As shown in Figure 23, there were three threshold parameter estimates with four response categories, (1) Never, (2) Sometimes, (3) Often, and (4) Very often, for the Quantitative Reasoning EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the QRproblem item had the highest threshold parameter value at all $b$ levels ($b_1 = -0.85, b_2 = 0.23, b_3 = 1.16$), indicating that of the Quantitative Reasoning EI items, the QRproblem item had a lower likelihood to be endorsed at the higher response value for each response category threshold. The standardized local dependence statistics indicated that all items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values $> 10$, ranging from 26.4 to 43.9. All of the local dependence values for the Quantitative Reasoning EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair.
The category response curves and information functions for the Quantitative Reasoning engagement indicator items are shown in Figure 23 below. For all three items, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta value where there was a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative items (indicated by the dashed information function line) were QRproblem and QRevaluate. The total information curve for the Quantitative Reasoning engagement indicator showed that the QR EI items could best be used from theta values of approximately -1.5 to 1.5 (See Figure 24). This indicated that the Quantitative Reasoning engagement indicator scores are least discriminating and informative for theta values below -1.5 and above 1.5 (i.e., students who reported the lowest (i.e., below -1.5 theta) and most frequent (i.e., above 1.5 theta) amounts of engagement in quantitative reasoning practices).

Figure 23

Category Response Curves and Information Functions for Quantitative Reasoning EI

Note. Category response curves are indicated by solid lines and the information functions are shown by dotted lines.
Figure 24

*Total Information Curve for Quantitative Reasoning EI*

Unidimensional IRT of Learning Strategies Engagement Indicator. The results of the unidimensional IRT analysis for the Learning Strategies engagement indicator indicated goodness of fit based on the $\chi^2(51) = 269.25, p < .01$ and the RMSEA value of 0.05 and $G^2(41) = 231.26, p < .01$ and the RMSEA values of 0.06. Both the Pearson $\chi^2$ and likelihood ratio statistic $G^2$ tests of model fit were statistically significant which could indicate poor fit; note that this result also may be attributed to the large sample size. Item parameter estimates indicated that the most discriminating item in the Learning Strategies EI was LSsummary ($a = 4.34$), while the least discriminating item was LSreading ($a = 1.35$). As shown in Figure 25, there were three threshold parameter estimates with four response categories, (1) Never, (2) Sometimes, (3) Often, and (4) Very often, for the Learning Strategies EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level
categories across items, the LSsummary item had the highest threshold parameter value at the $b_1$ level ($b_1 = -1.48$), indicating that of the Learning Strategies EI items, the LSsummary item had a lower likelihood to be endorsed at the higher response value, “Sometimes.” At the $b_3$ level, LSnotes and LSsummary ($b_3 = 0.75$) had lower likelihoods to be endorsed at the higher response value, “Very often.” The standardized local dependence statistics indicated that all items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values $> 10$, ranging from 10.7 to 28.2. All of the local dependence values for the Learning Strategies EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair.

The category response curves and information functions for the Learning Strategies engagement indicator items are shown in Figure 25 below. For items, excluding LSreading, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta value where there was a higher probability for each category to be endorsed than other response categories. The LSreading item had one response category, “Never,” which did not have a higher probability for being endorsed over other response categories; and, therefore, was a poorly functioning response option. As illustrated by the height of the dotted lines, the most informative item (indicated by the dashed information function line) was LSsummary. In contrast, LSreading had a very low information function, which was another indicator of the item functioning poorly. The total information curve for the Learning Strategies engagement indicator showed that the LS EI items could best be used from theta values of approximately -2 to 1.5 (See Figure 26). This indicated that the Learning Strategies engagement indicator scores were least discriminating and informative for theta values below -2 and above 1.5 (i.e., students who
reported the lowest (i.e., below -2 theta) and most frequent (i.e., above 1.5 theta) amounts of engagement in learning strategies).

**Figure 25**

*Category Response Curves and Information Functions for Learning Strategies EI*

Note. Category response curves are indicated by solid lines and the information functions are shown by dotted lines.

**Figure 26**

*Total Information Curve for Learning Strategies EI*

**MIRT of Academic Challenge Theme.** The multidimensional IRT analysis goodness of fit Pearson $\chi^2$ statistics for the Academic Challenge theme (comprised of the Higher Order Learning, Reflective and Integrative Learning, Quantitative Reasoning, and Learning Strategies
EI items) could not be computed. The underlying contingency table was too sparse to compute the full-information fit values because the total number of cross-classifications was $4^{17}$, which is vastly larger than the sample size of 1430. The $G^2(1255) = 27561.68$, $p < .01$ and the RMSEA value greater than .08 did not indicate acceptable fit; however, the $G^2$ values should be interpreted with caution given the issue with the sparse contingency table. Item parameter estimates indicated that the most discriminating item in the Academic Challenge theme was HOanalyze ($a = 4.10$), while the least discriminating item was LSreading ($a = 1.53$). The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the QRproblem item had the highest threshold parameter value at all $b$ levels ($b_1 = -0.84$, $b_2 = 0.22$, $b_3 = 1.16$), indicating that of the Academic Challenge theme items, the QRproblem item had a lower likelihood to be endorsed at the higher response value for each response category threshold.

The item parameter estimates were similar for the unidimensional analyses of the Academic Challenge themes’ four engagement indicators and for the multidimensional analysis of the Academic Challenge theme. The standardized local dependence statistics indicated that the approximately half of the item pairs were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values > 10, ranging from 1.0 to 42.8. Approximately 60% of the local dependence values for the Academic Challenge theme were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair. The local independence violations for the items within their
engagement indicator were similar to those in the unidimensional analyses; however, the multidimensional IRT analyses resulted in far more items that met the local independence assumption with items that were not in their engagement indicator.

**Learning with Peers Theme**

**Unidimensional IRT of Collaborative Learning Engagement Indicator.** The results of the unidimensional IRT analysis for the Collaborative Learning engagement indicator indicated goodness of fit based on the $\chi^2(239) = 877.97, p < .01$ and the RMSEA value of 0.04 and the $G^2(137) = 566.21, p < .01$ and the RMSEA value of 0.05. Both the Pearson $\chi^2$ and likelihood ratio statistic $G^2$ tests of model fit were statistically significant which could indicate poor fit; note that this result also may be attributed to the large sample size. Item parameter estimates indicated that the most discriminating item in the Collaborative Learning EI was CLstudy ($a = 3.74$), while the least discriminating items were CLexplain and CLproject ($a = 1.78$). As shown in Figure 27, there were three threshold parameter estimates with four response categories, (1) Never, (2) Sometimes, (3) Often, and (4) Very often, for the Learning Strategies EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the CLstudy item had the highest threshold parameter value at the $b_1$ level ($b_1 = -1.23$), indicating that of the Collaborative Learning EI items, the CLstudy item had a lower likelihood to be endorsed at the higher response value, “Sometimes.” At the $b_3$ level, CLexplain ($b_3 = 1.10$) had a lower likelihood to be endorsed at the higher response value, “Very often.” The standardized local dependence statistics indicated that there were numerous violations of the local independence assumption (i.e., an examinee’s response to different items
in an instrument are independent when the trait level is held constant) with absolute values > 10, ranging from 4.5 to 35.4. Two-thirds of the local dependence values for the Collaborative Learning EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

The category response curves and information functions for the Collaborative Learning engagement indicator items are shown in Figure 27 below. For all three items, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta value where there was a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative item (indicated by the dashed information function line) was CLstudy. The total information curve for the Collaborative Learning engagement indicator showed that the CL EI items could best be used from theta values of approximately -2.5 to 1.75 (See Figure 28). This indicated that the Collaborative Learning engagement indicator scores are least discriminating and informative for theta values above 1.75 (i.e., students who reported the most engagement in collaborative learning strategies).
Figure 27

*Category Response Curves and Information Functions for Collaborative Learning EI*

![Category Response Curves](image1)

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.

Figure 28

*Total Information Curve for Collaborative Learning EI*

![Total Information Curve](image2)

**Unidimensional IRT of Discussions with Diverse Others Engagement Indicator.** The results of the unidimensional IRT analysis for the Discussions with Diverse Others engagement indicator did not indicate goodness of fit based on the $\chi^2(239)= 4492.38$, $p < .01$ and the RMSEA value greater than .08; while the $G^2(106) = 1065.35$, $p < .01$ and the RMSEA value of 0.08 did fall within an acceptable value. Both the Pearson $\chi^2$ and likelihood ratio statistic $G^2$ tests of
model fit were statistically significant which could indicate poor fit; note that this result also may be attributed to the large sample size. Item parameter estimates indicated that the most discriminating item in the Discussions with Diverse Others EI was DDreligion \((a = 4.24)\), while the least discriminating item was DDrace \((a = 3.29)\). As shown in Figure 29, there were three threshold parameter estimates with four response categories, (1) Never, (2) Sometimes, (3) Often, and (4) Very often, for the Discussions with Diverse Others EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the \(b\) values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the DDpolitical item had the highest threshold parameter value at all \(b\) levels \((b_1 = -1.48, b_2 = -0.48, b_3 = 0.29)\), indicating that of the Discussions with Diverse Others EI items, the DDpolitical item had a lower likelihood to be endorsed at the higher response value for each response category threshold. The standardized local dependence statistics indicated that all items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values > 10, ranging from 40.3 to 60.6. Two-thirds of the local dependence values for the Discussions with Diverse Others EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

The category response curves and information functions for the Discussions with Diverse Others engagement indicator items are shown in Figure 29 below. For all four items, each response category (indicated by the four solid lines) worked distinctively, as shown by each
having a theta value where there was a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative item (indicated by the dashed information function line) was DDreligion. The total information curve for the Discussions with Diverse Others EI showed that the DD EI items could best be used from theta values of approximately -2.5 to 0.75 (See Figure 30). This indicated that the Discussions with Diverse Others EI scores were least discriminating and informative for theta values below -2.5 and above 0.75 (i.e., students who reported the lowest (i.e., below -2.5 theta) and more frequent (i.e., above 0.75 theta) amounts of engagement in discussions with diverse others).

**Figure 29**

*Category Response Curves and Information Functions for Discussion with Diverse Others EI*

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.
MIRT of Learning with Peers Theme. The multidimensional IRT analysis goodness of fit Pearson $\chi^2$ statistics for the Learning with Peers theme (comprised of the Collaborative Learning and Discussion with Diverse Others EI items) could not be computed. The underlying contingency table was too sparse to compute the full-information fit values because the total number of cross-classifications was $4^8 = 65,536$, which is vastly larger than the sample size of 1430. The $G^2(788)= 4971.46$, $p < .01$ and the RMSEA value of 0.06; note that the results indicated that the $G^2$ values should be interpreted with caution given the issue with the sparse contingency table. Item parameter estimates indicated that the most discriminating item in the Learning with Peers Theme was DDreligion ($a = 4.14$), while the least discriminating item was CLproject ($a = 1.79$). The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the CLstudy item had the highest threshold parameter value at the $b_1$ and
$b_2$ levels ($b_1 = -1.24$, $b_2 = -0.07$). At the $b_3$ level, CLexplain had the highest threshold parameter value ($b_3 = 1.09$). This indicated that of the Learning with Peers Theme items, for the $b_1$ and $b_2$ response category thresholds, the CLstudy item, and for the $b_3$ response category, the CLexplain item, had lower likelihoods to be endorsed at the higher response value for each response category threshold.

The item parameter estimates were similar for the unidimensional analyses of the Learning with Peers themes’ two engagement indicators and for the multidimensional analysis of the Learning with Peers theme. The standardized local dependence statistics indicated that approximately 1/3 of the item pairs were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values $> 10$, ranging from 1.8 to 59.7. Approximately half of the local dependence values for the Learning with Peers theme were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair. The local independence violations for the items were all within their engagement indicator, similar to those in the unidimensional analyses. The multidimensional IRT analyses resulted in all item pairs for items in different engagement indicators meeting the local independence assumption.

**Experiences with Faculty Theme**

**Unidimensional IRT of Student-Faculty Interactions Engagement Indicator.** The results of the unidimensional IRT analysis for the Student-Faculty Interactions engagement indicator indicated goodness of fit based on the $\chi^2(239) = 859.41$, $p < .01$ and the RMSEA value of 0.04 and $G^2(147) = 497.93$, $p < .01$ and the RMSEA value of 0.04. Both the Pearson $\chi^2$ and
likelihood ratio statistic $G^2$ tests of model fit were statistically significant which could indicate poor fit; note that this result also may be attributed to the large sample size. Item parameter estimates indicated that the most discriminating item in the Student-Faculty Interactions EI was SFdiscuss ($a = 3.10$), while the least discriminating items were SFcareer and SFotherwork ($a = 2.35$). As shown in Figure 31, there were three threshold parameter estimates with four response categories, (1) Never, (2) Sometimes, (3) Often, and (4) Very often, for the Student-Faculty Interactions EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the SFotherwork item had the highest threshold parameter value at all $b$ levels ($b_1 = -0.20$, $b_2 = 0.76$, $b_3 = 1.66$), indicating that of the Student-Faculty Interactions EI items, the SFotherwork item had a lower likelihood to be endorsed at the higher response value for each response category threshold. The standardized local dependence statistics indicated that the majority of items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values $> 10$, ranging from 5.0 to 20.8. One-third of the local dependence values for the Student-Faculty Interactions EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

The category response curves and information functions for the Student-Faculty Interactions engagement indicator items are shown in Figure 31 below. For all four items, each response category (indicated by the four solid lines) worked distinctively, as shown by each
having a theta value where there is a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative item (indicated by the dashed information function line) was SFdiscuss. The total information curve for the Student-Faculty Interactions EI showed that the SF EI items could best be used from theta values of approximately -1.5 to 2.5 (See Figure 32). This indicated that the Student-Faculty Interactions EI scores were least discriminating and informative for theta values below -1.5 and above 2.5 (i.e., students who reported the lower (i.e., below -1.5 theta) and most frequent (i.e., above 2.5 theta) amounts of interaction with faculty).

**Figure 31**

*Category Response Curves and Information Functions for Student-Faculty Interactions EI*

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.
Unidimensional IRT of Effective Teaching Practices Engagement Indicator. The results of the unidimensional IRT analysis for the Effective Teaching Practices engagement indicator did not indicate goodness of fit based on the $\chi^2(1003) = 13861.45, p < .01$ and the RMSEA value greater than .08; while the $G^2(245) = 1030.47, p < .01$ and the RMSEA value of 0.05 did fall within an acceptable value. Both the Pearson $\chi^2$ and likelihood ratio statistic $G^2$ tests of model fit were statistically significant which could indicate poor fit; note that this result also may be attributed to the large sample size. Item parameter estimates indicated that the most discriminating item in the Effective Teaching Practices EI was ETexample ($a = 2.91$), while the least discriminating item was ETdraftfb ($a = 1.99$). As shown in Figure 33, there were three threshold parameter estimates with four response categories, (1) Very little, (2) Some, (3) Quite a bit, and (4) Very much, for the Effective Teaching Practices EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not
indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the ETfeedback item had the highest threshold parameter value at all b levels (b₁ = -1.73, b₂ = -0.40, b₃ = 0.87), indicating that of the Effective Teaching Practices EI items, the ETfeedback item had a lower likelihood to be endorsed at the higher response value for each response category threshold. The standardized local dependence statistics indicated that the majority of items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values > 10, ranging from 4.6 to 34.4. Approximately two-thirds of the local dependence values for the Effective Teaching Practices EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

The category response curves and information functions for the Effective Teaching Practices engagement indicator items are shown in Figure 33 below. For all five items, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta value where there is a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative item (indicated by the dashed information function line) was ETexample. The total information curve for the Effective Teaching Practices EI showed that the ET EI items could best be used from theta values of approximately -3 to 1.5 (See Figure 34). This indicated that the Effective Teaching Practices EI scores were least discriminating and informative for theta values above 1.5 (i.e., students who reported the most frequent amounts of effective teaching practices used by instructors).
Figure 33

Category Response Curves and Information Functions for Effective Teaching Practices EI

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.

Figure 34

Total Information Curve for Effective Teaching Practices EI
**MIRT of Experiences with Faculty Theme.** The multidimensional IRT analysis goodness of fit Pearson $\chi^2$ statistics for the Experiences with Faculty theme (comprised of the Student-Faculty Interactions and Effective Teaching Practices EI items) could not be computed. The underlying contingency table was too sparse to compute the full-information fit values because the total number of cross-classifications was $4^9 = 266,144$, which is vastly larger than the sample size of 1430. The $G^2(999) = 6868.68$, $p < .01$ and the RMSEA value of 0.06 did fall within an acceptable value; however, the $G^2$ values should be interpreted with caution given the issue with the sparse contingency table. Item parameter estimates indicated that the most discriminating item in the Experiences with Faculty Theme was SFdiscuss ($\alpha = 3.06$), while the least discriminating item was ETdraftfb ($\alpha = 2.04$). The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the SFotherwork item had the highest threshold parameter value at all the $b$ levels ($b_1 = -0.20$, $b_2 = 0.76$, $b_3 = 1.66$), which indicated that of the Experiences with Faculty Theme items, the SFotherwork item had a lower likelihood to be endorsed at the higher response value for each response category threshold.

The item parameter estimates were similar for the unidimensional analyses of the Experiences with Faculty themes’ two engagement indicators and for the multidimensional analysis of the Experiences with Faculty theme. The standardized local dependence statistics indicated that approximately 1/3 of the item pairs were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values $> 10$, ranging from 1.4 to 32.7. Half of the local dependence values for the Experiences with Faculty theme were negative, indicating
that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair. The local independence violations for the items were all within their engagement indicator, similar to those in the unidimensional analyses. The multidimensional IRT analyses resulted in all item pairs for items in different engagement indicators meeting the local independence assumption.

**Campus Environment Theme**

**Unidimensional IRT of Quality of Interactions Engagement Indicator.** The unidimensional IRT analysis goodness of fit Pearson $\chi^2$ statistics for the Quality of Interactions engagement indicator could not be computed because the underlying contingency table was too sparse to compute the full-information fit values. The total number of cross-classifications was $7^5 = 16807$, which is much larger than the sample size of 1430. The $G^2(666) = 5099.71, p < .01$ and the RMSEA value of 0.07 did fall within an acceptable value; however, the $G^2$ values should be interpreted with caution given the issue with the sparse contingency table. The standardized local dependence statistics indicated that all items, except QIstudent, were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values > 10, ranging from 4.4 to 29.7. Half of the local dependence values for the Quality of Interactions EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

Item parameter estimates indicated that the most discriminating item in the Quality of Interactions EI was QIadmin ($a = 3.26$), while the least discriminating item was QIstudent ($a =$
0.95). As shown in Figure 35, there were six threshold parameter estimates with seven response categories, (1) Poor to (7) Excellent, for the Quality of Interactions EI items. The seven response categories are 0 to 6 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the QIadmin item had the highest threshold parameter value at the $b_1$ to $b_5$ levels ($b_1 = -1.56, b_2 = -1.14, b_3 = -0.69, b_4 = -0.17, b_5 = 0.46$), indicating that of the Quality of Interactions EI items, the QIadmin item had a lower likelihood to be endorsed at the higher response value for the first five response category thresholds. At the $b_6$ level, the QIfaculty item had the highest threshold parameter level ($b_6 = 1.26$).

The category response curves and information functions for the Quality of Interactions engagement indicator items are shown in Figure 35 below. For two items, QIstaff and QIadmin, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta value where there is a higher probability for each category to be endorsed than other response categories. However, even though each had a distinctive category, the range of each was very narrow. For the three items QIstudent, QIadvisor, and QIfaculty, there were poorly functioning response categories which did not have a higher probability for being endorsed over other response categories. Overall, the Quality of Interactions seven Likert point scale did not function well. As illustrated by the height of the dotted lines, the most informative item (indicated by the dashed information function line) was QIadmin. In contrast, QIstudent had a very low information function, which was another indicator of the item functioning poorly; QIadvisor and QIfaculty also had lower information functions, but were better than the QIstudent item. The total information curve for the Quality of Interactions engagement indicator showed
that the QI EI items could best be used from theta values of approximately -2.5 to 2 (See Figure 36). This indicated that the Quality of Interactions engagement indicator scores were least discriminating and informative for theta values below -2.5 and above 2 (i.e., students who reported the lowest (i.e., below -2.5 theta) and highest (i.e., above 2 theta) quality of interactions with people in their learning environment).

**Figure 35**

*Category Response Curves and Information Functions for Quality of Interactions EI*

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.
Unidimensional IRT of Supportive Environment Engagement Indicator. The unidimensional IRT analysis goodness of fit Pearson $\chi^2$ statistics for the Supportive Environment engagement indicator could not be computed because the underlying contingency table was too sparse to compute the full-information fit values. The total number of cross-classifications was $4^8 = 65536$, which is much larger than the sample size of 1430. The $G^2(917) = 6766.86, p < .01$ and the RMSEA value of 0.07 did fall within an acceptable value; however, the $G^2$ values should be interpreted with caution given the issue with the sparse contingency table. Item parameter estimates indicated that the most discriminating item in the Supportive Environment EI was SEwellness ($a = 2.66$), while the least discriminating item was SElearnsup ($a = 1.78$). As shown in Figure 37, there were three threshold parameter estimates with four response categories, (1) Very little, (2) Some, (3) Quite a bit, and (4) Very much, for the Supportive Environment EI items. The four response categories are 0 to 3 in the figure. The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher.
value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the SEnonacad item had the highest threshold parameter value at all $b$ levels ($b_1 = -0.52$, $b_2 = 0.66$, $b_3 = 1.74$), indicating that of the Supportive Environment EI items, the SEnonacad item had a lower likelihood to be endorsed at the higher response value for each category threshold. The standardized local dependence statistics indicated that all items were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is held constant) with absolute values $> 10$, ranging from 10.7 to 48.9. Three-quarters of the local dependence values for the Supportive Environment EI were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair.

The category response curves and information functions for the Supportive Environment engagement indicator items are shown in Figure 37. For all eight items, each response category (indicated by the four solid lines) worked distinctively, as shown by each having a theta value where there is a higher probability for each category to be endorsed than other response categories. As illustrated by the height of the dotted lines, the most informative items (indicated by the dashed information function line) were SESocial and SEwellness. The total information curve for the Supportive Environment engagement indicator showed that the SE EI items could best be used from theta values of approximately -2 to 2 (See Figure 38). This indicated that the Supportive Environment engagement indicator scores were least discriminating and informative for theta values below -2 and above 2 (i.e., students who reported the lowest (i.e., below -2 theta)
and most frequent (i.e., above 2 theta) amounts of emphasis the university places on supportive environment practices).

**Figure 37**

*Category Response Curves and Information Functions for Supportive Environment EI*

*Note.* Category response curves are indicated by solid lines and the information functions are shown by dotted lines.

**Figure 38**

*Total Information Curve for Supportive Environment EI*
MIRT of Campus Environment Theme. The multidimensional IRT analysis goodness of fit Pearson $\chi^2$ statistics for the Campus Environment theme (comprised of the Quality of Interactions and Supportive Environment EI items) could not be computed. The underlying contingency table was too sparse to compute the full-information fit values because the total number of cross-classifications was $7^5 + 4^8 = 82,343$, which is vastly larger than the sample size of 1430. The $G^2(928) = 27254.25, p < .01$ and the RMSEA value greater than .08 did not indicate acceptable fit; note that the $G^2$ values should be interpreted with caution given the issue with the sparse contingency table. Item parameter estimates indicated that the most discriminating item in the Campus Environment Theme was QIadmin ($a = 3.15$), while the least discriminating item was QIstudent ($a = 1.00$). The item threshold parameter estimates indicated that the $b$ values for all items increased consistently from lower to higher value; therefore, it did not indicate that there were any problems with ordering of the response categories. Comparing the same level categories across items, the SEnonacad item had the highest threshold parameter value at the $b_1$ to $b_3$ levels ($b_1 = -0.51, b_2 = 0.66, b_3 = 1.73$), indicating that of the Campus Environment Theme items, the SEnonacad item had a lower likelihood to be endorsed at the higher response category for each response category threshold. The Quality of Interactions EI items had more response categories than the Supportive Environment EI items: at the $b_4$ to $b_6$ levels, the QIadmin item had the highest parameter threshold values ($b_4 = -0.17, b_5 = 0.47, b_6 = 1.26$).

The item parameter estimates were similar for the unidimensional analyses of the Campus Environment themes’ two engagement indicators and for the multidimensional analysis of the Campus Environment theme. The standardized local dependence statistics indicated that approximately half of the item pairs were in violation of the local independence assumption (i.e., an examinee’s response to different items in an instrument are independent when the trait level is
held constant) with absolute values > 10, ranging from 0.8 to 47.4. Approximately two-thirds of the local dependence values for the Campus Environment theme were negative, indicating that the model implied correlation was higher than the observed correlation for an item pair; positive local dependence values indicated that the model implied correlation was lower than the observed correlation for an item pair. The local independence violations for the items were all within their engagement indicator, similar to those in the unidimensional analyses; all of the Supportive Environment item pairs were in violation of the local independence assumption. The multidimensional IRT analyses resulted in all item pairs for items in different engagement indicators meeting the local independence assumption.

**Confirmatory Factor Analysis**

CFA analyses were conducted on each of the four NSSE theme, beginning with configural invariance analysis, followed by 1st order model, 2nd order model, and bifactor model testing.

**Academic Challenge Theme**

*Configural Invariance Analysis of Academic Challenge Theme.* For the Academic Challenge theme, the configural invariance analysis goodness of fit results indicated an adequate fit of the model to the data: $\chi^2(260) = 1212.273, p < .001; \text{RMSEA} = 0.07; \text{CFI} = 0.97; \text{and TLI} = 0.97$. The factor loadings for each group, freshmen and seniors, indicated similar factor loadings of items across the four engagement indicators that comprise the Academic Challenge theme: Higher Order Learning items (freshman $\lambda = 0.79$ to 0.88; seniors $\lambda = 0.80$ to 0.90), Reflective and Integrative Learning (freshman $\lambda = 0.65$ to 0.80; seniors $\lambda = 0.75$ to 0.83), Quantitative Reasoning (freshman $\lambda = 0.79$ to 0.85; seniors $\lambda = 0.81$ to 0.92), and Learning Strategies (freshman $\lambda = 0.78$ to 0.86; seniors $\lambda = 0.76$ to 0.84). These results indicated that there was
configural invariance for freshmen and seniors; therefore, a combined sample could be utilized for further analyses on the Academic Challenge theme items.

**CFA of Academic Challenge Theme.** The results of the CFA analysis for the seventeen items loading onto the four engagement indicators as proposed by the NSSE are presented in Figure 39. The model demonstrated acceptable goodness of fit: RMSEA = 0.08; CFI = 0.97, and TLI = 0.96. All items demonstrated moderate to strong positive loadings onto their engagement indicator: Higher-Order Learning ($\lambda = 0.80 – 0.89$), Reflective and Integrative Learning ($\lambda = 0.74 – 0.83$), Quantitative Reasoning ($\lambda = 0.80 – 0.90$), and Learning Strategies ($\lambda = 0.76 – 0.89$). The correlations among the engagement indicators ranged from 0.42 to 0.57 indicating a moderate degree of correlation among factors.
Figure 39

Academic Challenge 1st Order CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Quantitative Reasoning (QR), Learning Strategies (LS)

(RMSEA = .08; CFI = .97; TLI = .96)
2nd Order CFA of Academic Challenge Theme. The results of the 2nd order CFA analyses for the seventeen items loading onto the four indicators, and in turn loading onto the Academic Challenge theme (see Figure 40) demonstrated acceptable goodness of fit: RMSEA = 0.07; CFI = 0.97, and TLI = 0.97. All items demonstrated moderate to strong positive loadings onto their engagement indicator: Higher-Order Learning ($\lambda = 0.80 - 0.89$), Reflective and Integrative Learning ($\lambda = 0.74 - 0.83$), Quantitative Reasoning ($\lambda = 0.80 - 0.90$), and Learning Strategies ($\lambda = 0.76 - 0.89$). The engagement indicator loadings onto the 2nd order Academic Challenge theme were all moderate to strong: Higher-Order Learning ($\lambda = 0.75$), Reflective and Integrative Learning ($\lambda = 0.76$), Quantitative Reasoning ($\lambda = 0.63$), and Learning Strategies ($\lambda = 0.66$). The $R^2$ value for each of the engagement indicators were moderate, indicating that a moderate amount of the variation in the engagement indicators was explained by the 2nd order factor (Academic Challenge theme): Higher-Order Learning ($R^2 = 0.57$), Reflective and Integrative Learning ($R^2 = 0.58$), Quantitative Reasoning ($R^2 = 0.39$), and Learning Strategies ($R^2 = 0.43$).
Figure 40

Academic Challenge (AC) 2nd Order CFA Model: Reflective & Integrative Learning (RI), Higher-Order Learning (HO), Learning Strategies (LS), Quantitative Reasoning (QR)

(RMSEA = .07; CFI = .97; TLI = .97)
Bifactor CFA of Academic Challenge. The results of the bifactor CFA analyses for the seventeen items loading onto the four engagement indicators and also the general factor of the Academic Challenge theme (see Figure 41) demonstrated acceptable goodness of fit: RMSEA = 0.06; CFI = 0.98, and TLI = 0.98. Items demonstrated low to moderate loadings onto their engagement indicator: Higher-Order Learning ($\lambda = 0.49 – 0.67$), Reflective and Integrative Learning ($\lambda = 0.39 – 0.63$), Quantitative Reasoning ($\lambda = 0.64 – 0.71$), and Learning Strategies ($\lambda = 0.31 – 0.73$). The loading of the items onto the general Academic Challenge theme were also low to moderate ($\lambda = 0.44$ to 0.70). There was not a consistent pattern of stronger loadings on the engagement indicator or general factor across all items. The Higher-Order Learning items had either similar loadings between the general factor and the engagement indicator (HOapply and HOnalyze), or slightly higher loadings on the general factor (HOevaluate and HOform). The Reflective and Integrative Learning items did not have similar patterns of loadings across all items; there were similar loadings between the general factor and the engagement indicator (RI societal, RI ownview, and RI perspect), higher loadings on the general factor (RI integrate, RN newview, and RC connect), and also an item with a higher loading on the engagement indicator (RD diverse). The Quantitative Reasoning items all had higher loadings on the engagement indicator (QR conclude, QR problem, and QR evaluate). The Learning Strategies items did not have similar patterns of loadings across all items; one item had a much higher loading on the general factor (L S reading), while the other two items had higher loadings on the engagement indicator (L S notes and L S summary). The comparison of the $G$ factor and unique factor (engagement indicator) contributions are shown using the R-squared values in Table 8.
Figure 41

Academic Challenge Bifactor CFA Model: Higher-Order Learning (HO), Reflective & Integrative Learning (RI), Quantitative Reasoning (QR), and Learning Strategies (LS)

(RMSEA = .06; CFI = .98; TLI = .98)
Table 8

*G factor and Unique Factor Contributions in Academic Challenge Bifactor Model*

<table>
<thead>
<tr>
<th></th>
<th>HO</th>
<th>RI</th>
<th>QR</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Factor Contribution</td>
<td>0.32</td>
<td>0.27</td>
<td>0.46</td>
<td>0.37</td>
</tr>
<tr>
<td>G Factor Contribution</td>
<td>0.41</td>
<td>0.37</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td>Error</td>
<td>0.27</td>
<td>0.37</td>
<td>0.25</td>
<td>0.35</td>
</tr>
</tbody>
</table>

**Comparison of CFA Models: Academic Challenge Theme.** To compare the CFA models, a chi-square difference test analysis was conducted on the 1st order and 2nd order models. However, the chi-square difference test could not be used for comparison of the bifactor model with the 1st order or 2nd order models because they are non-nested models. The chi-square difference testing indicated that the difference between the 1st order and 2nd order model was not statistically significant: $\chi^2(2) = 4.709, p \geq .05)$. The CFI and TLI values for both models indicated goodness of fit (CFI and TLI $\geq .95$). In addition, the RMSEA value for both models indicated goodness of fit with RMSEA $\leq .08$. Therefore, it could not be concluded that the 2nd order model is a better fitting model than the 1st order model.

A comparison of the goodness of fit parameters for the 1st order, 2nd order, and bifactor models of the Academic Challenge theme indicated that the best fitting model was the bifactor model, with the next best fitting model the 2nd order model, followed by the 1st order model. The loading value distributions for the items onto their engagement indicator across the three models were Higher-Order Learning (1st and 2nd order $\lambda = 0.80 – 0.89$; bifactor $\lambda = 0.49 – 0.67$), Reflective and Integrative Learning (1st and 2nd order $\lambda = 0.74 – 0.83$; bifactor $\lambda = 0.39 – 0.63$), Quantitative Reasoning (1st and 2nd order $\lambda = 0.80 – 0.90$; bifactor $\lambda = 0.64 – 0.71$), and Learning Strategies (1st and 2nd order $\lambda = 0.76 – 0.89$; bifactor $\lambda = 0.31 – 0.73$). The 1st order model indicated a moderate degree of correlation among the engagement indicators. In the 2nd
order model, the engagement indicator loadings onto the 2nd order Academic Challenge theme were moderate to strong: Higher-Order Learning ($\lambda = 0.75$), Reflective and Integrative Learning ($\lambda = 0.76$), Quantitative Reasoning ($\lambda = 0.63$), and Learning Strategies ($\lambda = 0.66$). In the bifactor model, the loading of the items onto the general Academic Challenge theme were low to moderate ($\lambda = 0.44$ to $0.70$), with the contribution from the unique factors (i.e., engagement indicators) and general factor (i.e., the theme) somewhat evenly distributed, and between $0.25$ to $0.37$ attributed to error (as shown in Table 8). The RMSEA value was lowest (best) for the bifactor model (0.06), then the 2nd order model (0.07), followed by the 1st order model (0.08). The CFI and TLI values were highest (best) for the bifactor model (CFI = 0.98; TLI = 0.98), then the 2nd order model (CFI = 0.97; TLI = 0.97), followed by the 1st order model (CFI = 0.97; TLI = 0.96).

**Learning with Peers Theme**

**Configural Invariance Analysis of Learning with Peers Theme.** For the Learning with Peers theme, the configural invariance analysis goodness of fit results indicated an adequate fit of the model to the data: $\chi^2 (54) = 266.398, p < .001$; RMSEA = 0.07; CFI = 0.99; and TLI = 0.99. The factor loadings for each group, freshmen and seniors, indicated similar factor loadings of items across the two engagement indicators that comprise the Learning with Peers theme: Collaborative Learning (freshman $\lambda = 0.66$ to $0.87$; seniors $\lambda = 0.71$ to $0.88$) and Discussions with Diverse Others (freshman $\lambda = 0.89$ to $0.93$; seniors $\lambda = 0.87$ to $0.92$). These results indicated that there was configural invariance for freshmen and seniors; therefore, a combined sample could be utilized for further analyses on the Learning with Peers theme items.

**CFA of Learning with Peers Theme.** The results of the CFA analysis for the eight items loading onto the two engagement indicators as proposed by the NSSE are presented in Figure 42. The model demonstrated acceptable goodness of fit as the CFI and TLI values were excellent (CFI
= 0.99; TLI = 0.99), though the RMSEA was 0.085, slightly greater than 0.08. All items demonstrated moderate to strong positive loadings onto their engagement indicator: Collaborative Learning ($\lambda = 0.70 \text{ – } 0.88$) and Discussions with Diverse Others ($\lambda = 0.88 \text{ – } 0.91$). The correlation among the engagement indicators was 0.26 indicating a relatively low degree of correlation.

**Figure 42**

*Learning with Peers CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD)*

---

2nd Order CFA of Learning with Peers Theme. The Learning with Peers 2nd order CFA model did not converge; therefore, no results could be reported. Due to the inclusion of only two factors, the factor loadings from the Learning with Peers theme to its engagement indicators were specified with some restrictions, for example, by fixing the 2nd order loading to a...
constant in order to have an identified model. However, the models with different specifications did not converge due to a non-positive definite residual covariance matrix.

**Bifactor CFA of Learning with Peers Theme.** The results of the bifactor CFA analyses for the eight items loading onto the two engagement indicators and also the general factor of the Learning with Peers theme (see Figure 43) demonstrated acceptable goodness of fit: RMSEA = 0.06; CFI = 1.0, and TLI = 0.99. The Collaborative Learning items demonstrated moderate to strong loadings onto their engagement indicator ($\lambda = 0.66 – 0.86$). On the other hand, the Discussions with Diverse Others items had low negative loadings to moderate positive loadings on their engagement indicator ($\lambda = -0.17 – 0.43$). The loading of the items onto the general Academic Challenge theme were low for the Collaborative Learning items ($\lambda = 0.14 – 0.24$) and were strong for the Discussions with Diverse Others items ($\lambda = 0.83 – 0.96$). All of the Collaborative Learning items had a much higher loading on the engagement indicator; while all of the Discussions with Diverse Others items had a much higher loading on the general factor (Learning with Peers theme). The comparison of the $G$ factor and unique factor (engagement indicator) contributions are shown using the R-squared values in Table 9.
**Figure 43**

*Learning with Peers (LP) Bifactor CFA Model: Collaborative Learning (CL) and Discussions with Diverse Others (DD)*

![Diagram of Learning with Peers Bifactor CFA Model](image)

(RMSEA = .06; CFI = 1.0; TLI = .99)

**Table 9**

*Learning with Peers Bifactor Model Comparison of G factor and Unique Factor Contributions*

<table>
<thead>
<tr>
<th></th>
<th>CL</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Factor</td>
<td>0.54</td>
<td>0.07</td>
</tr>
<tr>
<td>Contribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G Factor Contribution</td>
<td>0.04</td>
<td>0.80</td>
</tr>
<tr>
<td>Error</td>
<td>0.42</td>
<td>0.13</td>
</tr>
</tbody>
</table>

**Comparison of CFA Models: Learning with Peers Theme.** The 1st order and bifactor models of the Learning with Peers theme were compared using goodness of fit parameters; the chi-square difference test could not be used for comparison because they are non-nested models.
The 2\textsuperscript{nd} order model for the Learning with Peers theme did not converge, so was eliminated from the possible models. The comparison of the goodness of fit parameters for the 1\textsuperscript{st} order and bifactor models of the Learning with Peers theme indicated that the best fitting model was the bifactor model. The loading value distributions for the items onto their engagement indicator across the two items were Collaborative Learning (1\textsuperscript{st} order $\lambda = 0.70 – 0.88$; bifactor $\lambda = 0.66 – 0.86$) and Discussions with Diverse Others (1\textsuperscript{st} order $\lambda = 0.88 – 0.91$; bifactor $\lambda = -0.17 – 0.43$). The 1\textsuperscript{st} order model indicated a relatively low degree of correlation among the engagement indicators. In the bifactor model, the loading of the items onto the general Learning with Peers theme were low for the Collaborative Learning items ($\lambda = 0.14 – 0.24$) and strong for the Discussions with Diverse Others items ($\lambda = 0.83 – 0.96$). The unique factor contribution was moderate at 0.54 for Collaborative Learning with only 0.04 contributed to the general factor (i.e., Learning with Peers theme), while 0.42 was attributed to error; in contrast, the unique factor contribution was very low at 0.07 for Discussions with Diverse Others with 0.80 contributed to the general factor, and 0.13 attributed to error (as shown in Table 9). The RMSEA value was lowest (best) for the bifactor model (0.06), while the RMSEA value for the 1\textsuperscript{st} order model (0.09). The CFI and TLI values were highest (best) for the bifactor model (CFI = 1.0; TLI = 0.99), followed by the 1\textsuperscript{st} order model (CFI = 0.99; TLI = 0.99).

**Experiences with Faculty Theme**

**Configural Invariance Analysis of Experiences with Faculty Theme.** For the Experiences with Faculty theme, the configural invariance analysis goodness of fit results indicated an adequate fit of the model to the data: $\chi^2(70) = 453.621, p < .001$; RMSEA = 0.09; CFI = 0.97; and TLI = 0.97. The RMSEA was slightly higher than the recommended value of $\leq 0.08$; however, at 0.088, and given the other goodness of fit values, the model fit was deemed
acceptable. The factor loadings for each group, freshmen and seniors, indicated similar factor loadings of items across the two engagement indicators that comprise the Experiences with Faculty theme: Student-Faculty Interaction (freshman $\lambda = 0.73$ to 0.83; seniors $\lambda = 0.79$ to 0.86) and Effective Teaching Practices (freshman $\lambda = 0.75$ to 0.82; seniors $\lambda = 0.79$ to 0.85). These results indicated that there was configural invariance for freshmen and seniors; therefore, a combined sample could be utilized for further analyses on the Experiences with Faculty theme items.

**CFA of Experiences with Faculty Theme.** The results of the CFA analysis for the nine items loading onto the two engagement indicators as proposed by the NSSE are presented in Figure 44. The model did not demonstrate satisfactory goodness of fit as the RMSEA was greater than 0.08 (RMSEA = .10); however, the CFI and TLI values were good (CFI = 0.98; TLI = 0.97). All items demonstrated moderate to strong loadings onto their engagement indicator: Student-Faculty Interactions ($\lambda = 0.79$ – 0.85) and Effective Teaching Practices ($\lambda = 0.78$ – 0.84). The correlation among the engagement indicators was 0.31 indicating a relatively low degree of correlation.
2nd Order CFA of Experiences with Faculty Theme. The Experiences with Faculty 2nd order CFA model did not converge; therefore, no results could be reported. Due to the inclusion of only two factors, the factor loadings from the Experiences with Faculty theme to its engagement indicators were specified with some restrictions, for example, by fixing the 2nd order loading to a constant in order to have an identified model. However, the models with different specifications did not converge due to a non-positive definite residual covariance matrix.

Bifactor CFA of Experiences with Faculty Theme. The results of the bifactor CFA analyses for the nine items loading onto the two engagement indicators and also the general factor for the Experiences with Faculty theme (see Figure 45) demonstrated acceptable goodness
of fit: RMSEA = 0.04; CFI = 1.0, and TLI = 0.99. The Student-Faculty Interaction items demonstrated strong loadings onto their engagement indicator ($\lambda = 0.72 – 0.81$). The Effective Teaching Practices items had moderate to strong loadings on their engagement indicator ($\lambda = 0.56 – 0.87$). The loading of the items onto the general Experiences with Faculty theme were low to moderate for the Student-Faculty Interaction items ($\lambda = 0.25 – 0.35$) and were moderate to strong for the Effective Teaching Practices items ($\lambda = 0.47 – 0.87$). All of the Student-Faculty Interaction items had a much higher loading on the engagement indicator than the general factor. Four of the Effective Teaching Practices items had similar loadings on the engagement indicator and general factor (ETgoals, ETexample, ETdraftfb, and ETfeedback); while one item had a much higher loading on the engagement indicator (ETorganize). The comparison of the $G$ factor and unique factor (engagement indicator) contributions are shown using the R-square values in Table 10.
Figure 45

Experiences with Faculty (EF) Bifactor CFA Model: Student-Faculty Interactions (SF) and Effective Teaching Practices (ET)

Table 10

Experiences with Faculty Bifactor Model Comparison of G factor and Unique Factor Contributions

<table>
<thead>
<tr>
<th></th>
<th>SF</th>
<th>ET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique Factor Contribution</td>
<td>0.57</td>
<td>0.29</td>
</tr>
<tr>
<td>G Factor Contribution</td>
<td>0.10</td>
<td>0.45</td>
</tr>
<tr>
<td>Error</td>
<td>0.34</td>
<td>0.26</td>
</tr>
</tbody>
</table>
Comparison of CFA Models: Experiences with Faculty Theme. The 1st order and bifactor models of the Experiences with Faculty theme were compared using goodness of fit parameters; the chi-square difference test could not be used for comparison because they are non-nested models. The 2nd order model for the Experiences with Faculty theme did not converge, so was eliminated from the possible models. The comparison of the goodness of fit parameters for the 1st order and bifactor models of the Experiences with Faculty theme indicated that the best fitting model was the bifactor model. The loading value distributions for the items onto their engagement indicator across the two items were Student-Faculty Interactions (1st order $\lambda = 0.79 – 0.85$; bifactor $\lambda = 0.72 – 0.81$) and Effective Teaching Practices (1st order $\lambda = 0.78 – 0.84$; bifactor $\lambda = 0.56 – 0.87$). The 1st order model indicated a relatively low degree of correlation among the engagement indicators. In the bifactor model, the loading of the items onto the general Experiences with Faculty theme were low to moderate for the Student-Faculty Interaction items ($\lambda = 0.25 – 0.35$) and were moderate to strong positive for the Effective Teaching Practices items ($\lambda = 0.47 – 0.87$). The unique factor contribution was moderate at 0.57 for Student-Faculty Interactions with only 0.10 contributed to the general factor (i.e., Experiences with Faculty theme), while 0.34 was attributed to error; in contrast, the unique factor contribution was low at 0.29 for Effective Teaching Practices with 0.45 contributed to the general factor, and 0.26 attributed to error (as shown in Table 10). The RMSEA value was lowest (best) for the bifactor model (0.04), while the RMSEA value for the 1st order model (0.10). The CFI and TLI values were highest (best) for the bifactor model (CFI = 1.0; TLI = 0.99), followed by the 1st order model (CFI = 0.98; TLI = 0.97).
**Campus Environment Theme**

**Configural Invariance Analysis of Campus Environment Theme.** For the Campus Environment theme, the configural invariance analysis goodness of fit results indicated an adequate fit of the model to the data: $\chi^2(169) = 848.353$, $p < .001$; RMSEA = 0.08; CFI = 0.96; and TLI = 0.97. The factor loadings for each group, freshmen and seniors, indicated similar factor loadings of items across the two engagement indicators that comprise the Campus Environment theme: Quality of Interactions (freshman $\lambda = 0.55$ to 0.85; seniors $\lambda = 0.56$ to 0.88) and Supportive Environment (freshman $\lambda = 0.70$ to 0.81; seniors $\lambda = 0.71$ to 0.81). These results indicated that there was configural invariance for freshmen and seniors; therefore, a combined sample could be utilized for further analyses on the Campus Environment theme items.

**CFA of Campus Environment Theme.** The results of the CFA analyses for the thirteen items loading onto the two engagement indicators proposed by the NSSE are presented in Figure 46. The model did not demonstrate goodness of fit as the RMSEA was greater than 0.08 (RMSEA = 0.093); however, the CFI and TLI values were acceptable (CFI = 0.96; TLI = 0.95). Items demonstrated moderate to strong loadings onto their engagement indicator: Quality of Interactions ($\lambda = 0.55 – 0.85$) and Supportive Environment ($\lambda = 0.72 – 0.81$). The item with the lowest factor loading was QIstudent ($\lambda = 0.55$), which also had a lower factor loading in the EFA results. The correlation among the engagement indicators was 0.50 indicating a moderate degree of correlation among factors.
Figure 46

Campus Environment CFA Model: Quality of Interactions (QI) and Supportive Environment (SE)

(QMSEA = .09; CFI = .96; TLI = .95)
2nd Order CFA of Campus Environment Theme. The Campus Environment 2nd order CFA model did not converge; therefore, no results could be reported. Due to the inclusion of only two factors, the factor loadings from the Campus Environment theme to its engagement indicators were specified with some restrictions, for example, by fixing the 2nd order loading to a constant in order to have an identified model. However, the models with different specifications did not converge due to a non-positive definite residual covariance matrix.

Bifactor CFA of Campus Environment Theme. The Campus Environment bifactor model did not converge; therefore, no results could be reported. The model was attempted using the different estimator choices available within MPlus (WLS, WLSM, WLSMV, ML, MLR, MLF, and ULS); however, the model would not converge with any of the estimator options.

Comparison of CFA Models: Campus Environment Theme. For the Campus Environment theme, the 2nd order and bifactor models did not converge, so comparison of goodness of fit across models was not possible. While the 1st order model converged, it did not indicate goodness of fit. The RMSEA value was greater than 0.08 (RMSEA = 0.093); however, the CFI and TLI values were acceptable (CFI = 0.96; TLI = 0.95).
Chapter 5: Discussion

The National Survey of Student Engagement (NSSE) was developed in the hopes of providing valid and reliable information about institutional quality and student engagement, and is administered to hundreds of thousands of freshmen and seniors each year at over 500 institutions. This study addressed the dearth of research investigating the psychometric properties and validity evidence of the revised NSSE 2.0, including the validity of the proposed conceptual structure for measuring student engagement. Research has shown that changes to the content of a measurement instrument can have significant influences on the psychometric properties of the instrument, and therefore when an instrument undergoes major revisions, its psychometric properties must be reexamined to ensure that the uses and interpretation of scores are psychometrically appropriate (Furr, 2011).

Through a review of the literature, two NSSE 2.0 studies on national data (Miller et al., 2016; Zilvinskis et al., 2017) found evidence of the validity of the instrument. Miller et al. (2016), used EFA and CFA and found evidence supporting the validity of the engagement indicator and theme model. Zilvinskis et al. (2017) used canonical correlation analysis to examine the relationship between NSSE scores and self-reported learning gains of the NSSE and NSSE 2.0, and found convergent validity evidence for both NSSE versions, but only discriminant validity evidence for NSSE 2.0. In contrast, two validation studies of NSSE 2.0 (Le, 2019; Winkler, 2020) on single institution’s data raised concerns about the reliability and validity of the NSSE’s ten engagement indicators and four themes on their data sets. However, no research was identified that investigated whether the proposed NSSE 2.0 ten engagement indicator and four theme model was corroborated through higher-order factor analysis. In addition, no research was found that utilized IRT to analyze the psychometric properties of the
NSSE 2.0 items. Only one study was found that examined the pre-2013 version of the NSSE utilizing IRT; before the instrument underwent a significant revision. The current study addressed the research gap through the following research questions: (a) Based on unidimensional item response theory (IRT) and multidimensional item response theory (MIRT) criteria, are the NSSE items of the ten engagement indicators psychometrically sound?; (b) Based on confirmatory factor analysis (CFA) criteria, does the NSSE’s suggested factor structure of the construct of student engagement with ten engagement indicators organized into four themes fit the UHM 2015 NSSE data, and does a bifactor or second-order factor structure display a better fit?; and (c) What do the integrated results of (a) and (b) reveal about the psychometric adequacy of the NSSE 2.0? The findings in relation to each of the research questions are discussed in the subsequent sections.

**Exploratory Factor Analysis**

The EFA analysis on all 47 NSSE 2.0 items together resulted in a ten-factor solution, which corresponded with the ten engagement indicator groupings proposed by the NSSE. In addition, the four EFA analyses of the NSSE 2.0 items by each theme (i.e., Academic Challenge, Learning with Peers, Experiences with Faculty, and Campus Environment) resulted in factor extractions that corresponded with the NSSE 2.0 engagement indicator groupings. These EFA results provided support for the NSSE 2.0 conceptual structure utilizing ten engagement indicators. Also, of note, the two items with the lowest loadings in the EFA analyses, LSreading and QIstudent, were identified in the IRT analyses as poor performing items.
**Item Response Theory**

**Unidimensional IRT**

The threshold parameter estimates increasing consistently indicated that the ordering of the response categories for all items was working well. The category response curves and information functions for 43 of the 47 items indicated that the items and their response categories were working well. This provided evidence that the majority of the NSSE 2.0 individual items (i.e., 43 of 47) worked well psychometrically. Four items were identified as poorly functioning with low discrimination values, low information function lines, and poorly functioning response categories. These four items had response categories which did not have a higher probability for being endorsed over other response categories. The poorly functioning response categories were the LSreading item’s lowest response category “Never;” the QIstudent item’s lowest three response categories “1,” “2,” and “3” on the (1) Poor to (7) Excellent scale; and the QIadvisor and QIfaculty items’ response category “3” on the 1 to 7 scale. The LSreading item asks respondents how often they identified key information from reading assignments; considering the importance of reading within college curriculum, the poor functioning of the “Never” category is not surprising. The Quality of Interactions items ask respondents to indicate the quality of their interactions with people at their institution (i.e., students, academic advisors, faculty). The poor functioning of the lowest three categories for student to student interactions in the QIstudent item indicated that overall those interactions are more positive, so the lowest rating options are less often endorsed. For academic advisor to student (i.e., QIadvisor) and faculty to student (i.e., QIfaculty) interactions, the middle category was less often endorsed, indicating that students were more likely to endorse towards the negative or positive ends of the scale, and not
the neutral “3” response. Therefore, at the individual item level, only four of the items had item level statistics that indicated they were not psychometrically sound.

The total information curves for the ten engagement indicators revealed that across most of the engagement indicators, the items were least discriminating and informative for the students who reported the highest levels of engagement (i.e., the highest response category options) in the behaviors and practices measured by the items. This can be explained by the fact that the “most engaged” students were predicted to respond to each item in the highest response category, and, therefore, cannot be differentiated from each other. In an IRT study of another higher education survey, Sharkness and DeAngelo (2011) also found that items were least discriminating for the students with the highest involvement levels because those students were predicated to all respond the same way, by selecting the highest category.

The unidimensional IRT analyses revealed violations of local independence for all ten engagement indicators and goodness of fit issues. The $G^2$ fit index resulted in acceptable goodness of fit for all ten of the NSSE engagement indicators. However, using the $\chi^2$ fit statistic, only three of the ten engagement indicators (Learning Strategies, Collaborative Learning, and Student-Faculty Interactions) demonstrated acceptable fit: the $\chi^2$ fit statistic is sensitive to larger sample size ($n = 1430$), which could have attributed to fewer engagement indicators meeting this goodness of fit criteria. For five of the engagement indicators (Higher-Order Learning, Quantitative Reasoning, Learning Strategies, Discussions with Diverse Others, and Supportive Environment), all item pairs demonstrated local independence violations. For the remaining five engagement indicators (Reflective and Integrative Learning, Collaborative Learning, Student-Faculty Interactions, Effective Teaching Practices, and Quality of Interactions), the majority of items violated the local independence assumption. The violations of local independence mean
that the items within the engagement indicator scales did not meet the IRT modeling assumption that the only influence on an individual’s item response is that of the latent trait variable being measured and that item responses are independent from other item responses, i.e., not being influenced by another item’s response (Toland, 2014), and it can distort the estimated item parameter (Sharkness & DeAngelo, 2011). The finding of local independence violations contradicts the claim by the NSSE developers that the ten engagement indicators are unidimensional constructs (The Center for Postsecondary Research, n.d.). While the NSSE developers have stated that IRT was utilized in the psychometric evaluation and development of the NSSE 2.0, there are no IRT studies available on the NSSE Psychometric Portfolio webpages, so examination of their data and findings to back the unidimensionality claim was not possible. The local dependence found among the items within the engagement indicators indicated that there was a more complicated structure to the NSSE 2.0 engagement indicators, suggesting the next step of the multidimensional IRT.

**Multidimensional IRT**

The MIRT GRM were better models overall than the unidimensional IRT models. The MIRT analyses revealed improved local independence, with fewer violations; therefore, correlation among the engagement indicators explained some of the variance not explained by unidimensional IRT. The improved MIRT models for the NSSE 2.0 data aligned with the research findings that when unidimensional IRT models are applied to multidimensional data, this usually results in biased parameter estimates (Ackerman, 1989; Ansley & Forsyth, 1985; Folk & Green, 1989; Way et al., 1988). MIRT provides more accurate estimates since it can account for correlations among subscales (de la Torre, 2008; de la Torre & Patz, 2005; de la Torre & Song, 2009; Sheng & Wikle, 2007; Wang et al., 2004; Yao & Boughton, 2007). For
three of the four NSSE themes (Learning with Peers, Experiences with Faculty, and Campus Environment), all of the local independence violations occurred within engagement indicators (i.e., among items within the same engagement indicator). For the Academic Challenge theme, most of the local independence violations were within engagement indicators.

The \( \chi^2 \) fit statistic could not be computed for any of the four themes because the underlying contingency tables were too sparse to compute the full-information fit values; this was due to the total number of cross-classifications being much larger than the sample size. The Learning with Peers and Experiences with Faculty themes had acceptable fit based on the \( G^2 \) fit index. The Academic Challenge and Campus Environment themes did not have acceptable \( G^2 \) fit values; however, given the issue with the sparse contingency tables, the \( G^2 \) value results should be interpreted with caution. The item parameter estimates were similar for the unidimensional analyses of the engagement indicators and for the MIRT of each theme. The improvement in the models when the engagement indicators were grouped by theme in the four MIRT analyses, versus the unidimensional IRT engagement indicator models, provided partial confirmation that the four NSSE 2.0 themes are indeed the overall umbrella concepts for their corresponding engagement indicators. All four themes had improved local independence, and two of the themes also indicated goodness of fit (i.e., Learning with Peers and Experiences with Faculty). The lack of fit for the Academic Challenge theme may be related to the CFA finding that the best fit for this theme was actually a bifactor model.

**Confirmatory Factor Analysis**

Configural invariance analyses results indicated that there was configural invariance for freshmen and seniors for each of the four themes; therefore, a combined sample was utilized for further analyses. The CFA analyses for each theme indicated that the Academic Challenge and
Learning with Peers CFA models had acceptable goodness of fit, although the RMSEA value for the Learning with Peers CFA model was borderline acceptable at 0.85. In contrast, the Experiences with Faculty and Campus Environment CFA models did not have acceptable goodness of fit based on their RMSEA values. All four themes had acceptable CFI and TLI values. There was a moderate degree of correlation among engagement indicators within the Academic Challenge and Campus Environment themes; while the engagement indicators within the Learning with Peers and Experiences with Faculty themes had a relatively low degree of correlation. Based on the CFA model results, it could only be partially concluded that NSSE’s factor structure of ten engagement indicators fit the UHM data. The current study’s CFA findings correspond with the findings of Le (2019) whose CFA study of the NSSE 2.0 engagement indicator model for transfer students using a single institution’s data also found mixed results of model fit across the engagement indicator themes. For Le (2019), only the Academic Challenge theme indicated goodness of fit across the transfer subgroups examined.

Only the 2nd order model for the Academic Challenge theme converged; the 2nd order model for the other three themes were eliminated as possible models. The Learning with Peers, Experiences with Faculty, and Campus Environment themes only have two engagement indicators (i.e., factors) each, which caused a model identification issue; models with the engagement indicator factor loadings to the themes (i.e., the 2nd order factor) fixed at one, to allow for model identification, also resulted in non-convergence. Therefore, the NSSE themes could not be confirmed through higher-order CFA for the Learning with Peers, Experiences with Faculty, and Campus Environment themes. The Academic Challenge 2nd order model demonstrated acceptable goodness of fit with moderate to strong loadings of the items onto their engagement indicator, and moderate to strong loadings of the engagement indicators onto the
academic Challenge theme (i.e., the 2nd order factor). An analysis of a third-order factor structure was not possible due to the nonconvergence of the 2nd order models for three of the four NSSE 2.0 themes.

The bifactor models for the Academic Challenge, Learning with Peers, and Experiences with Faculty themes indicated acceptable goodness of fit. However, the bifactor model for the Campus Environment theme did not converge, and was, therefore, eliminated as a possible model. There was not a consistent pattern of stronger loadings on the engagement indicator or general factor (i.e., theme) across all items for the three bifactor models. For the Academic Challenge theme, the loadings onto the general (i.e., theme) and unique (i.e., engagement indicators) factors were somewhat evenly distributed with most items having moderate loadings on both, and between 0.25 to 0.37 attributed to error; therefore, for the Academic Challenge theme, a bifactor score interpretation utilizing both the engagement indicator and general Academic Challenge theme score would provide a reliable score interpretation. The engagement indicator scores would reflect information above and beyond the latent trait reflected in the general score, Academic Challenge. However, for the Learning with Peers and Experiences with Faculty themes, the loadings onto unique factors (i.e., engagement indicators) and general factors (i.e., themes) were not consistent within each model which makes the application of a bifactor score approach unreliable. For the Learning with Peers theme, the unique factor contribution was moderate at 0.54 for Collaborative Learning with only 0.04 contributed to the general factor (i.e., Learning with Peers theme) and 0.42 was attributed to error; in contrast, the unique factor contribution was very low at 0.07 for Discussions with Diverse Others with 0.80 contributed to the general factor, and 0.13 attributed to error. For the Experiences with Faculty theme, the unique factor contribution was moderate at 0.57 for Student-Faculty Interactions with only 0.10
contributed to the general factor (i.e., Experiences with Faculty theme), while 0.34 was attributed to error; in contrast, the unique factor contribution was low at 0.29 for Effective Teaching Practices with 0.45 contributed to the general factor, and 0.26 attributed to error.

While a comparison of model fit utilizing the chi-square difference test analysis is ideal, it requires that models be nested, and the bifactor models were not nested models with the CFA or 2nd order models. Therefore, the chi-square difference test analysis could only be utilized for the comparison of the Academic Challenge theme’s 1st order and 2nd order CFA models. The chi-square difference test for these two models was not statistically significant, so it could not be concluded that the 2nd order model was better fitting than the CFA model for the Academic Challenge theme. The bifactor models were compared with the 1st and 2nd order models for the Academic Challenge theme, and with the CFA models for the Learning with Peers and Experiences with Faculty themes utilizing the goodness of fit statistics. For all three themes, the bifactor model was identified as the best fitting model based on the goodness of fit statistics.

While the goodness of fit indices suggested that the bifactor models had the best fit, an inspection of the factor loadings for each item onto their engagement indicator and the general factor (i.e., theme) did not show a consistent pattern of stronger loadings on the engagement indicator or general factor across all items. This inconsistency makes it difficult to interpret the general factor and its applicability and usefulness in NSSE score interpretation. According to DeMars (2013),

when loadings on the specific factors are high, factor score estimates for the specific scores can be meaningful as long as score users are carefully educated about the fact that each subscale reflects information above and beyond the skill or trait reflected in the
general score. When loadings on the specific factors are low, only the general factor score carries a reliable interpretation (p. 374-375).

For bifactor models to provide meaningful scoring inferences, there needs to be consistency in the loadings on the general and specific factors across items; this was only found in the UHM data for the Academic Challenge theme. While the bifactor models for Learning with Peers and Experiences with Faculty themes cannot be recommended for use in a bifactor scoring method of the NSSE 2.0, based on the UHM data, the results of the general and unique factor contributions for these two themes did show that there are two different factors in each model, thus supporting a two factor CFA model for each. Therefore, the bifactor results provide support for the conceptual structure of the NSSE 2.0 for the Academic Challenge, Learning with Peers, and Experiences with Faculty themes.

Overall, the CFA (1st order, 2nd order, and bifactor) results provide partial support for the separate scoring and use of individual engagement indicator scores, while also providing evidence that there is a multidimensional nature to the NSSE 2.0 engagement indicators. The results provide support for the use of engagement indicator scores in the Academic Challenge (i.e., Higher-Order Learning, Reflective and Integrative Learning, Quantitative Reasoning, and Learning Strategies) and Learning with Peers themes (i.e., Collaborative Learning and Discussions with Diverse Others). The results for the Experiences with Faculty theme indicated that the engagement indicators do comprise two factors; however, due to the issues with goodness of fit of the 1st order CFA model and the inconsistency in loadings of the bifactor model, it could not be definitively concluded that the Student-Faculty Interactions and Effective Teaching Practices engagement indicator scores can be used as individual scores. Similarly, for the Campus Environment theme, the 1st order CFA results did not indicate goodness of fit, and it
could not be concluded that the Quality of Interactions and Supportive Environment engagement indicators can be used as separate scores; however, the moderate correlation between these engagement indicators may suggest that, despite not being able to directly test through 2nd order and bifactor models, the grouping of these two engagement indicators together may be warranted.

Of note is that for the current study, the RMSEA value of 0.08 or less was utilized as the criteria for reasonable fit. This contrasts with the RMSEA value of 0.06 criteria that the NSSE developers utilized in their CFA NSSE 2.0 construct validity study published on the NSSE Psychometric Portfolio webpage (The Center for Postsecondary Research, 2013). If the NSSE criteria of RMSEA of 0.06 or less had been utilized, this would have further reduced the models that indicated acceptable fit to only include the three bifactor models for the Academic Challenge, Learning with Peers, and Experiences with Faculty themes. All the 1st order CFA models and the Academic Challenge 2nd order model that had indicated acceptable fit using the 0.08 criteria would have been eliminated as possible models.

**Integration of IRT and CFA Results**

The results of the IRT and CFA analyses both indicated that there was a multidimensional or higher order nature to the NSSE engagement indicators. The unidimensional IRT analyses revealed significant violations of local independence; likewise, the CFA models did not indicate that all of the ten NSSE engagement indicators fit the UHM data. The IRT analyses added valuable information on item level characteristics that allowed for the identification of four poorly functioning items (i.e., LSreading, QIstudent, QIadvisor, and QIfaculty). The MIRT analyses indicated that correlation among the engagement indicators explained some of the variance not explained by the unidimensional IRT. For the Academic
Challenge theme, both the 2nd order and bifactor models indicated goodness of fit. The fit statistics suggested that the bifactor model had the best fit for the Academic Challenge, Learning with Peers, and Experiences with Faculty themes, which also indicated that a multidimensional model may better capture the engagement indicator data.

These findings align with the theoretical framework of the study which conceptualized student engagement as a multidimensional construct drawing upon two key theories of student engagement: Fredricks et al.’s (2004) tripartite model of student engagement and Kuh et al.’s (2007b) framework of student success. Drawing upon Kuh et al.’s framework of student success, the complexity of the multidimensional construct of student engagement becomes clear. The NSSE was designed to measure behavioral and cognitive dimensions, not emotional dimensions as found in Fredricks et al.’s tripartite model. There are numerous antecedents; interrelated influences of the cognitive, emotional, and behavioral dimensions of student engagement; and consequences of engagement. This kind of complexity is difficult to capture within a survey instrument due to limitations in survey length and respondent survey fatigue.

The Campus Environment theme did not indicate goodness of fit in the MIRT or CFA analyses. Therefore, the data clearly indicated issues with the Campus Environment theme whose item’s assess students’ perceived quality of interactions with important individuals in their learning environment (i.e., students, advisors, faculty, staff, and administrators) and the amount of emphasis the institution placed on programs and practices that support student persistence and learning (i.e., academic support programs, learning support services, encouraging contact among diverse students, opportunities for social involvement, campus activities, health and wellness, and support for non-academic responsibilities).
Psychometric Quality and Intended Use of the NSSE

In Fosnacht and Gonyea’s (2018) generalizability study of the NSSE 2.0 engagement indicators utilizing national data, they found that only some of the engagement indicators could be dependably generalized to a higher-order construct when using small samples. The authors reported that “when the object of measurement is an institution, the indicators with lower levels of dependability when generalizing over students and items would be most reliably treated as indexes (groups of items that, when combined, indicate a more general characteristic) rather than higher-order constructs” (Fosnacht & Gonyea, 2018, p. 69). Their finding aligns with the mixed results of the current study. Similarly, Winkler (2020) utilized generalizability theory to examine the NSSE 2.0 engagement indicators and found that while the engagement indicator scores were reliable at the individual student level, they “were not sufficiently reliable nor substantively meaningful to warrant use for institutional decision-making” at the college level (p. iii).

The NSSE is intended to be used for institutional decision making about policies and practices related to undergraduate student experiences in and out of the classroom, and also for benchmarking with other institutions (The Center for Postsecondary Research, 2017). Fosnacht and Gonyea (2018) argued that the psychometric issues found with the engagement indicators as higher order constructs is expected because the items were not selected at random from a pool of all the possible questions related to each higher order construct, but rather function as an index or snapshot of the level of engagement in specific beneficial activities known to improve university outcomes” (p. 70). However, regardless of the item selection process utilized by the NSSE staff, for institutions to be able to make confident decisions about and changes to institutional policies and practice, the measurement quality of the NSSE 2.0 needs to be accurate. The findings of the current study partially support that the NSSE 2.0 has adequate psychometric quality to make
confident decisions utilizing the NSSE engagement indicator scores at the institution level. The Academic Challenge and Learning with Peers themes were deemed adequate psychometrically for their intended use; however, the findings for the Experiences with Faculty theme were inconclusive and the Campus Environment theme was lacking in psychometric quality. At the item level, the majority of the engagement indicator items appeared to be working well based on item parameter estimates and information functions, though there was local dependency among items of the engagement indicators. Both the CFA and MIRT results indicated that there was a multidimensional or higher order nature to the NSSE engagement indicators. However, due to mixed results across engagement indicators and themes and issues with goodness of fit of numerous models, the NSSE’s suggested model structure of the construct of student engagement with ten engagement indicators organized into four themes was not fully confirmed.

Recommendations

Recommendations for Future Research

Further IRT and MIRT analyses on NSSE 2.0 data samples are needed on national and institution level data sets. Most of the research investigating the NSSE has utilized principles based on classical test theory (CTT); however, item response theory can provide much richer analyses of measurement precision than can be achieved with CTT (Embretson & Reise, 2000; Sharkness & DeAngelo, 2011). In addition, MIRT provides better estimation when there are large numbers of dimensions and when there are a small number of items in each dimension, as found in the NSSE 2.0. Further investigation on additional data samples into the response category functioning for the four poorly functioning items (i.e., LSreading, QIstudent, QIadvisor, and QIfaculty) should examine if the identified response categories work distinctively: LSreading item’s lowest response category “Never;” the QIstudent item’s lowest three response
categories “1,” “2,” and “3” on the (1) Poor to (7) Excellent scale; and the QIadvisor and QIfaculty items’ response category “3” on the 1 to 7 scale.

In the current study, the findings indicated a multidimensional nature to the NSSE 2.0 engagement indicators and, based on the CFA model fit statistics, the bifactor models appeared to have the best fit. However, based on the loadings of the general (i.e., theme) and specific (i.e., engagement indicator) factors, the utilization of a bifactor scoring approach was only found to be applicable to the Academic Challenge theme. For the Learning with Peers and Experiences with Faculty themes, the inconsistency in loading patterns onto the general and specific factors made the practical application of the bifactor models to scoring have no meaningful use. To gain a better understanding of the construct structure of the NSSE 2.0, it would be beneficial for further studies to be conducted utilizing multidimensional models.

Future studies on larger samples across institutions and administration years of the NSSE 2.0 will aid in determining if the psychometric issues identified in this study reach beyond a single institution data set. Stronger evidence is needed to confirm the recommended use of the engagement indicator scores in the Experiences with Faculty and Campus Environment themes, warranting future and further studies. At the UHM, the NSSE 2.0 was administered for the first time since the 2015 administration in the 2020-21 academic year, a follow-up study to determine if the same issues are present in the newer survey administration sample should be conducted.

**Recommendations for Policy and Practice**

The findings also warrant recommendations in relation to policy and practice. Based on the findings of inconclusive and inadequate psychometric quality for the Experiences with Faculty and Campus Environment themes in the current study, I recommend that university administrators triangulate their institution’s NSSE 2.0 results with additional data sources prior
to the allocation of resources and/or changes to policy and practice based on NSSE results for these themes. In addition, based on the IRT results for the four poorly functioning items (i.e., LSreading, QIstudent, QIadvisor, and QIfaculty), these items may be improved by a modification to the response categories and/or the collapsing of the poor performing response categories for score calculation. The three Learning Strategies items utilize a four-point frequency scale starting with “Never” to ask students how often they identify key material in a reading assignment, review their class notes, and summarize what they’ve learned in class. As changing the response scale for only one of the three items would be problematic for score calculation, a modification to begin the scale with a low frequency, such as “Rarely,” instead of “Never,” may improve the functioning of the LSreading item. In addition, the Quality of Interaction items may be significantly improved by reducing the response categories to a four-point or five-point scale, from the current seven-point scale, and based on the poor functioning of the middle (i.e., neutral) “3” category, a four-point scale may be found superior to a five-point scale.

Finally, while the NSSE psychometric portfolio webpage states that NSSE staff designed the engagement indicators to be unidimensional constructs and that they utilized IRT in the testing of the NSSE 2.0, their IRT results are not published in the NSSE Psychometric Portfolio. In contrast, the NSSE staff have published their results using other methodologies, such as EFA, CFA, and predictive validity (The Center for Postsecondary Research, n.d.). It would be beneficial for the institutional researchers and administrators using the NSSE 2.0 at their institutions to be able to examine IRT and MIRT results conducted on the national sample of NSSE 2.0 data. I recommend that the NSSE staff publish IRT studies in the NSSE Psychometric Portfolio.
**Limitations of the Study**

There are several limitations to this study. First, the data were from a single institution and single administration year. Second, there were issues with the nonconvergence of multiple models, which raises the question of whether the nonconvergence was sample specific or if the models would converge using larger cross institution and/or multi-year administration samples. It is possible that the nonconvergence was caused by the underlying theory of the conceptual structure itself or by the individual items (i.e., non-functioning categories). Finally, higher order IRT was not included in the analyses due to convergence issues with the higher order models, so a comparison of all the CFA models with MIRT models was not conducted.
References


Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. SAGE.


Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of educational Measurement, 47*(3), 361-372.


The Center for Postsecondary Research. (2019). *NSSE Annual Results 2019*. Indiana University School of Education


Appendix A: NSSE 2015, The College Student Report

1. During the current school year, about how often have you done the following?
   Response options: Very often, Often, Sometimes, Never
   a. Asked questions or contributed to course discussions in other ways
   b. Prepared two or more drafts of a paper or assignment before turning it in
   c. Come to class without completing readings or assignments
   d. Attended an art exhibit, play, or other arts performance (dance, music, etc.)
   e. Asked another student to help you understand course material
   f. Explained course material to one or more students
   g. Prepared for exams by discussing or working through course material with other students
   h. Worked with other students on course projects or assignments
   i. Given a course presentation

2. During the current school year, about how often have you done the following?
   Response options: Very often, Often, Sometimes, Never
   a. Combined ideas from different courses when completing assignments
   b. Connected your learning to societal problems or issues
   c. Included diverse perspectives (political, religious, racial/ethnic, gender, etc.) in course discussions or assignments
   d. Examined the strengths and weaknesses of your own views on a topic or issue
   e. Tried to better understand someone else’s views by imagining how an issue looks from his or her perspective
   f. Learned something that changed the way you understand an issue or concept
   g. Connected ideas from your courses to your prior experiences and knowledge

3. During the current school year, about how often have you done the following?
   Response options: Very often, Often, Sometimes, Never
   a. Talked about career plans with a faculty member
   b. Worked with a faculty member on activities other than coursework (committees, student groups, etc.)
   c. Discussed course topics, ideas, or concepts with a faculty member outside of class
   d. Discussed your academic performance with a faculty member

4. During the current school year, how much has your coursework emphasized the following?
   Response options: Very much, Quite a bit, Some, Very little
   a. Memorizing course material
   b. Applying facts, theories, or methods to practical problems or new situations
   c. Analyzing an idea, experience, or line of reasoning in depth by examining its parts
   d. Evaluating a point of view, decision, or information source
   e. Forming a new idea or understanding from various pieces of information

5. During the current school year, to what extent have your instructors done the following?
   Response options: Very much, Quite a bit, Some, Very little
   a. Clearly explained course goals and requirements
   b. Taught course sessions in an organized way
   c. Used examples or illustrations to explain difficult points
   d. Provided feedback on a draft or work in progress
   e. Provided prompt and detailed feedback on tests or completed assignments

6. During the current school year, about how often have you done the following?
   Response options: Very often, Often, Sometimes, Never
   a. Reached conclusions based on your own analysis of numerical information (numbers, graphs, statistics, etc.)
   b. Used numerical information to examine a real-world problem or issue (unemployment, climate change, public health, etc.)
   c. Evaluated what others have concluded from numerical information

Copyright © 2015 Trustees of Indiana University
Use of this survey without permission is prohibited.
National Survey of Student Engagement

7. During the current school year, about how many papers, reports, or other writing tasks of the following lengths have you been assigned? [Include those not yet completed.]
   Response options: None, 1-2, 3-5, 6-10, 11-15, 16-20, More than 20 papers
   a. Up to 5 pages
   b. Between 6 and 10 pages
   c. 11 pages or more

8. During the current school year, about how often have you had discussions with people from the following groups?
   Response options: Very often, Often, Sometimes, Never
   a. People of a race or ethnicity other than your own
   b. People from an economic background other than your own
   c. People with religious beliefs other than your own
   d. People with political views other than your own

9. During the current school year, about how often have you done the following?
   Response options: Very often, Often, Sometimes, Never
   a. Identified key information from reading assignments
   b. Reviewed your notes after class
   c. Summarized what you learned in class or from course materials

10. During the current school year, to what extent have your courses challenged you to do your best work?
    Response options: 1=Not at all to 7=Very much

11. Which of the following have you done or do you plan to do before you graduate?
    Response options: Done or in progress, Plan to do, Do not plan to do, Have not decided
    a. Participated in an internship, co-op, field experience, student teaching, or clinical placement
    b. Held a formal leadership role in a student organization or group
    c. Participated in a learning community or some other formal program where groups of students take two or more classes together
    d. Participated in a study abroad program
    e. Worked with a faculty member on a research project
    f. Completed a culminating senior experience (capstone course, senior project or thesis, comprehensive exam, portfolio, etc.)

12. About how many of your courses at this institution have included a community-based project [service-learning]?
    Response options: All, Most, Some, None

13. Indicate the quality of your interactions with the following people at your institution.
    Response options: 1=Poor to 7=Excellent, Not Applicable
    a. Students
    b. Academic advisors
    c. Faculty
    d. Student services staff (career services, student activities, housing, etc.)
    e. Other administrative staff and offices (Registrar, financial aid, etc.)

14. How much does your institution emphasize the following?
    Response options: Very much, Quite a bit, Some, Very little
    a. Spending significant amounts of time studying and on academic work
    b. Providing support to help students succeed academically
    c. Using learning support services (tutoring services, writing center, etc.)
    d. Encouraging contact among students from different backgrounds (social, racial/ethnic, religious, etc.)
    e. Providing opportunities to be involved socially
    f. Providing support for your overall well-being (recreation, health care, counseling, etc.)
    g. Helping you manage your non-academic responsibilities (work, family, etc.)
    h. Attending campus activities and events (performing arts, athletic events, etc.)
    i. Attending events that address important social, economic, or political issues
15. About how many hours do you spend in a typical 7-day week doing the following?  
   Response options: 0, 1-5, 6-10, 11-15, 16-20, 21-25, 26-30, More than 30 (Hours per week)  
   a. Preparing for class (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing, and other academic activities)  
   b. Participating in co-curricular activities (organizations, campus publications, student government, fraternity or sorority, intercollegiate or intramural sports, etc.)  
   c. Working for pay on campus  
   d. Working for pay off campus  
   e. Doing community service or volunteer work  
   f. Relaxing and socializing (time with friends, video games, TV or videos, keeping up with friends online, etc.)  
   g. Providing care for dependents (children, parents, etc.)  
   h. Commuting to campus (driving, walking, etc.)

16. Of the time you spend preparing for class in a typical 7-day week, about how much is on assigned reading?  
   Response options: Very little, Some, About half, Most, Almost all

17. How much has your experience at this institution contributed to your knowledge, skills, and personal development in the following areas?  
   Response options: Very much, Quite a bit, Some, Very little  
   a. Writing clearly and effectively  
   b. Speaking clearly and effectively  
   c. Thinking critically and analytically  
   d. Analyzing numerical and statistical information  
   e. Acquiring job- or work-related knowledge and skills  
   f. Working effectively with others  
   g. Developing or clarifying a personal code of values and ethics  
   h. Understanding people of other backgrounds (economic, racial/ethnic, political, religious, nationality, etc.)  
   i. Solving complex real-world problems  
   j. Being an informed and active citizen

18. How would you evaluate your entire educational experience at this institution?  
   Response options: Excellent, Good, Fair, Poor

19. If you could start over again, would you go to the same institution you are now attending?  
   Response options: Definitely yes, Probably yes, Probably no, Definitely no

20a. How many majors do you plan to complete? [Do not count minors.]  
   Response options: One, More than one

20b. [If answered “One”] Please enter your major or expected major: [Text box]

20c. [If answered “More than one”] Please enter up to two majors or expected majors (do not enter minors): [Text box]

21. What is your class level?  
   Response options: Freshman/first-year, Sophomore, Junior, Senior, Unclassified

22. Thinking about this current academic term, are you a full-time student?  
   Response options: Yes, No

23a. How many courses are you taking for credit this current academic term?  
   Response options: 0, 1, 2, 3, 4, 5, 6, 7 or more

23b. Of these, how many are entirely online?  
   Response options: 0, 1, 2, 3, 4, 5, 6, 7 or more

24. What have most of your grades been up to now at this institution?  
   Response options: A, A-, B+, B, B-, C+, C, C- or lower

Copyright © 2015 Trustees of Indiana University  
Use of this survey without permission is prohibited.
25. Did you begin college at this institution or elsewhere?
   Response options: Started here, Started elsewhere

26. Since graduating from high school, which of the following types of schools have you attended other than the one you are now attending? (Select all that apply.)
   Response options: Vocational or technical school, Community or junior college, 4-year college or university other than this one, None, Other

27. What is the highest level of education you ever expect to complete?
   Response options: Some college but less than a bachelor’s degree, Bachelor’s degree (B.A., B.S., etc.), Master’s degree (M.A., M.S., etc.), Doctoral or professional degree (Ph.D., J.D., M.D., etc.)

28. What is the highest level of education completed by either of your parents or (those who raised you)?
   Response options: Did not finish high school, High school diploma or G.E.D., Attended college but did not complete degree, Associate’s degree (A.A., A.S., etc.), Bachelor’s degree (B.A., B.S., etc.), Master’s degree (M.A., M.S., etc.), Doctoral or professional degree (Ph.D., J.D., M.D., etc.)

29. What is your gender identity?
   Response options: Man, Woman; Another gender identity, please specify; I prefer not to respond

30. Enter your year of birth (e.g., 1994):

31a. Are you an international student?
   Response options: Yes, No

31b. [If answered “yes”] What is your country of citizenship?

32. What is your racial or ethnic identification? (Select all that apply.)
   Response options: American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, White, Other, I prefer not to respond

33. Are you a member of a social fraternity or sorority?
   Response options: Yes, No

34. Which of the following best describes where you are living while attending college?
   Response options: Dormitory or other campus housing (not fraternity or sorority house), Fraternity or sorority house, Residence (house, apartment, etc.) within walking distance to the institution, residence (house, apartment, etc.) farther than walking distance to the institution, None of the above

35. Are you a student-athlete on a team sponsored by your institution’s athletics department?
   Response options: Yes, No

36. Are you a current or former member of the U.S. Armed Forces, Reserves, or National Guard?
   Response options: Yes, No

37a. Have you been diagnosed with any disability or impairment?
   Response options: Yes, No, I prefer not to respond

37b. [If answered “yes”] Which of the following has been diagnosed? (Select all that apply.)
   Response options: A sensory impairment (vision or hearing), A mobility impairment, A learning disability (e.g., ADHD, dyslexia), A mental health disorder, A disability or impairment not listed above

38. Which of the following best describes your sexual orientation? [Question administered per institution request.]
   Response options: Heterosexual, Gay, Lesbian, Bisexual, Another sexual orientation, please specify; Questioning or unsure, I prefer not to respond
Appendix B: NSSE 2015, Engagement Indicators and Items

Engagement Indicators and Items

Academic Challenge

Higher-Order Learning
During the current school year, how much has your coursework emphasized the following:
• Applying facts, theories, or methods to practical problems or new situations
• Analyzing an idea, experience, or line of reasoning in depth by examining its parts
• Evaluating a point of view, decision, or information source
• Forming a new idea or understanding from various pieces of information

Reflective & Integrative Learning
During the current school year, how often have you:
• Combined ideas from different courses when completing assignments
• Connected your learning to societal problems or issues
• Included diverse perspectives (political, religious, racial/ethnic, gender, etc.) in course discussions or assignments
• Examined the strengths and weaknesses of your own views on a topic or issue
• Tried to better understand someone else’s views by imagining how an issue looks from his or her perspective
• Learned something that changed the way you understand an issue or concept
• Connected ideas from your courses to your prior experiences and knowledge

Learning Strategies
During the current school year, how often have you:
• Identified key information from reading assignments
• Reviewed your notes after class
• Summarized what you learned in class or from course materials

Quantitative Reasoning
During the current school year, how often have you:
• Reached conclusions based on your own analysis of numerical information (numbers, graphs, statistics, etc.)
• Used numerical information to examine a real-world problem or issue (unemployment, climate change, public health, etc.)
• Evaluated what others have concluded from numerical information

Learning with Peers

Collaborative Learning
During the current school year, how often have you:
• Asked another student to help you understand course material
• Explained course material to one or more students
• Prepared for exams by discussing or working through course material with other students
• Worked with other students on course projects or assignments

Discussions with Diverse Others
During the current school year, how often have you had discussions with people from the following groups:
• People from a race or ethnicity other than your own
• People from an economic background other than your own
• People with religious beliefs other than your own
• People with political views other than your own

Experiences with Faculty

Student-Faculty Interaction
During the current school year, how often have you:
• Talked about career plans with a faculty member
• Worked with a faculty member on activities other than coursework (committees, student groups, etc.)
• Discussed course topics, ideas, or concepts with a faculty member outside of class
• Discussed your academic performance with a faculty member

Effective Teaching Practices
During the current school year, to what extent have your instructors done the following:
• Clearly explained course goals and requirements
• Taught course sessions in an organized way
• Used examples or illustrations to explain difficult points
• Provided feedback on a draft or work in progress
• Provided prompt and detailed feedback on tests or completed assignments

Campus Environment

Quality of Interactions
Indicate the quality of your interactions with the following people at your institution:
• Students
• Academic advisors
• Faculty
• Student service staff (career services, student activities, housing, etc.)
• Other administrative staff and offices (registrars, financial aid, etc.)

Supportive Environment
How much does your institution emphasize the following:
• Providing support to help students succeed academically
• Using learning support services (tutoring services, writing center, etc.)
• Encouraging contact among students from different backgrounds (social, racial/ethnic, religious, etc.)
• Providing opportunities to be involved socially
• Providing support for your overall well-being (recreation, health care, counseling, etc.)
• Helping you manage your non-academic responsibilities (work, family, etc.)
• Attending campus activities and events (performing arts, athletic events, etc.)
• Attending events that address important social, economic, or political issues

High-Impact Practice Items

Which of the following have you done or do you plan to do before you graduate?
• Participate in a study abroad program
• Work with a faculty member on a research project
• Complete a culminating senior experience (capstone course, senior project or thesis, comprehensive exam, portfolio, etc.)

About how many of your courses at this institution have included a community-based project (service-learning)?