


Spring 2017

# Student engagement and post-college outcomes: A comparison of formative and reflective models

Courtney Sanders  
*James Madison University*

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

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Methods Commons](#), and the [Higher Education Commons](#)

---

## Recommended Citation

Sanders, Courtney, "Student engagement and post-college outcomes: A comparison of formative and reflective models" (2017). *Dissertations*. 146.  
<https://commons.lib.jmu.edu/diss201019/146>

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

Student Engagement and Post-College Outcomes:  
A Comparison of Formative and Reflective Models  
Courtney B. Sanders

A dissertation submitted to the Graduate Faculty of  
JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

Department of Graduate Psychology

May 2017

---

FACULTY COMMITTEE:

Committee Chair: John D. Hathcoat, Ph.D.

Committee Members:

Keston H. Fulcher, Ph.D.

S. Jeanne Horst, Ph.D.

## Acknowledgments

I must begin by thanking my advisor, John Hathcoat. You have been a great source of support throughout my doctoral studies. I have learned so much from working with you and I have appreciated your philosophical insights and your outside-of-the-box thinking. Your constructive feedback was invaluable in strengthening my dissertation and as your first doctoral student, I appreciated that we could learn together on this journey.

I am also incredibly thankful to the members of my committee: Keston Fulcher and Jeanne Horst. Keston, thank you for always helping me to keep the big picture in mind and for your comments and feedback, which strengthened my writing. I am also grateful for your support and encouragement throughout my time in the Ph.D. program. Jeanne, thank you for your patience, detailed feedback, and thoughtful comments throughout this process. Your insights improved my dissertation and it was a joy working with you.

Thanks to all the students, faculty, and staff in CARS. I appreciate the opportunity I had to work with so many of you and the impact you made on my graduate experience. I could not imagine going through the doctoral journey without the support from my graduate school friends who I spent more late nights and early mornings with than I can count: Aaron Myers, Thai Ong, Kristen Smith, and Scott Strickman. I am thankful for the support network you provided and I am grateful to know each of you as a colleague and a friend. I look forward to what the future holds for all of us and I can't wait to see all the things each of you will accomplish. I especially must thank you, Scott. We started this program the same week, took almost every class together, and shared an office for two years. Thank you for helping me maintain humor on challenging days, for the constant

debriefing sessions, and for taking numerous walks around the lake with me. You are a great friend and I am thankful we started and finished this journey together.

Finally, I would like to thank my close friends and family. Alesia, thank you for being a friend who is more like a sister. Your constant prayers and words of encouragement fueled me through this journey and I am forever grateful to you for your support. Diana, thank you for taking the time to visit me, constantly making me laugh, and giving me a needed break from countless projects and tasks. To my brothers, Grant and Miles, thank you for telling me ridiculous jokes and for keeping me grounded during this process. Mom and Dad, I appreciate the constant source of love and support you provided me during the most difficult parts of this journey. Thank you for encouraging me to work hard, live with integrity, and reminding me to glorify God in everything that I do.

## Table of Contents

<b>Acknowledgments</b> .....	ii
<b>List of Tables</b> .....	vii
<b>List of Figures</b> .....	viii
<b>Abstract</b> .....	ix
<b>Introduction</b> .....	1
Student Engagement as a Measure of Quality in Higher Education .....	1
Benson’s Validity Framework .....	2
What is Student Engagement?.....	3
Prior Psychometric Evaluations of Student Engagement.....	5
Is Student Engagement a Formative or a Reflective Latent Variable? .....	6
Purpose of the Current Study .....	9
<b>Literature Review</b> .....	11
Engagement Theory and Research .....	11
Defining student engagement .....	11
Theoretical conceptualizations of student engagement .....	14
Recent developments in student engagement research .....	17
Measuring Student Engagement.....	19
Student engagement instruments .....	19
College Student Experiences Questionnaire .....	20
National Survey of Student Engagement .....	22
Psychometric evaluations of scores from the NSSE.....	25
Beginning College Survey of Student Engagement.....	27
Community College Survey of Student Engagement .....	28
The Gallup-Purdue Index .....	30
Student Engagement and Related Outcomes .....	32
Learning and developmental outcomes .....	32
Post-college outcomes .....	33
Life satisfaction.....	33
Career earnings .....	34
Workplace engagement and satisfaction.....	35
Problems when Measuring Student Engagement.....	35

Reflective measurement modeling .....	38
Formative measurement modeling .....	39
Identification of formative measurement models .....	41
Criticisms of formative measurement .....	42
Choosing a reflective or a formative approach.....	44
Direction of causality .....	44
Interchangeability of indicators .....	45
Covariance of indicators .....	46
Antecedents and consequences of indicators .....	46
Modeling Student Engagement in the Current Study.....	47
Evaluation of the GPI-SEAS using Benson’s Framework and Jarvis et al.’s Criteria .....	47
The Current Study .....	51
Stage I: Item analysis and estimating reflective models using exploratory factor analysis .....	51
Stage II: Estimating reflective models using confirmatory factor analysis.....	52
Stage III: Estimating formative models .....	52
<b>Methods</b> .....	54
Measures.....	54
GPI Student Experiences and Attachment Scale (GPI-SEAS).....	54
Cantril Self-Anchoring Striving Scale (Cantril Scale) .....	54
GPI Workplace Satisfaction Item (GPI-WSI) .....	55
Participants .....	55
Data Collection Procedures .....	55
Data Analyses.....	56
Stage I: Item analysis and reflective exploratory factor analysis using the first independent sample .....	57
Stage II: Estimating reflective CFA models using second independent sample .....	59
Data screening and assumptions .....	59
Structural equation modeling.....	60
Model identification.....	61
Estimation method .....	62
Evaluation of model-data fit .....	63

Chi-square goodness-of-fit test ( $\chi^2$ ) .....	63
Standard root mean square residual (SRMR) .....	63
Root mean square error of approximation (RMSEA).....	64
Comparative fit index (CFI).....	64
Evaluation of model parameters .....	64
Nested model comparisons .....	65
Full structural model.....	66
Stage III: Estimating the formative models using the second independent sample ..	67
Model identification.....	67
Evaluation of model-data fit .....	68
<b>Results</b> .....	69
Stage I: Item Analysis and Estimating Reflective Models using Exploratory Factor Analysis .....	69
Research question 1 .....	69
Research question 2 .....	70
Stage II: Estimating Reflective Models using Confirmatory Factor Analysis .....	73
Research question 3 .....	73
Data screening and guidelines for evaluating model fit .....	74
Reflective models: Confirmatory factor analysis .....	75
Research question 4.....	79
Stage III: Estimating Formative Models .....	82
Research question 5 .....	82
Research question 6.....	89
<b>Discussion</b> .....	90
Stage I.....	91
Stage II.....	93
Stage III .....	95
Recommendations for the Use of the GPI-SEAS.....	96
Limitations of the Current Study and Directions for Future Research.....	98
Implications and Conclusions.....	103
References.....	128

## List of Tables

Table 1. A Framework for Assessing Reflective and Formative Models.....	105
Table 2. Alignment of the GPI-SEAS Items with Gallup-Purdue Categories and Theoretical Affective and Behavioral Categories.....	106
Table 3. Item Polychoric Correlations and Descriptive Statistics for Subsample 1 .....	107
Table 4. Item Polychoric Correlations and Descriptive Statistics for Subsample 2 .....	108
Table 5. Summary of Results across All Stages of Analyses .....	109
Table 6. Parallel Analysis Based on Minimum Rank Factor Analysis (PA-MRFA) of Polychoric Correlations .....	110
Table 7. Exploratory Factor Analysis Structure (Pattern) Coefficients for Subsample 1 .....	111
Table 8. Fit Statistics for the Nested Confirmatory Factor Analysis Models.....	112
Table 9. Unstandardized (Standardized) Factor Pattern Coefficients and Standard Error Variances for the Three-Factor Model.....	113
Table 10. Polychoric Correlation Residuals Greater than $ .15 $ for One- and Two-Factor CFA Models.....	114
Table 11. Fit Statistics for Full Structural Models.....	115
Table 12. Full Structural Model B Structure (Pattern) Coefficients, Error Variances, Disturbance Terms, and Latent Variable Variances .....	116
Table 13. Fit Statistics for Formative MIMIC Models .....	117
Table 14. Unstandardized (Standardized) Factor Pattern Coefficients and Variance Explained for One-Construct Formative Model .....	118



## List of Figures

Figure 1. Example of general reflective and formative models.....	119
Figure 2. One-factor confirmatory factor analysis model.....	120
Figure 3. Two-factor confirmatory factor analysis model.....	121
Figure 4. Three-factor confirmatory factor analysis model.....	122
Figure 5. Full structural model A.....	123
Figure 6. Full structural model B.....	124
Figure 7. One-latent construct formative MIMIC model.....	125
Figure 8. Two-latent construct formative MIMIC model.....	126
Figure 9. Three-latent construct formative MIMIC model.....	127

## **Abstract**

Student engagement is a complex construct that is thought to be related to positive outcomes during and after college. Previous research has defined engagement in diverse ways and there are inconsistencies in the models that are used to measure this construct. Many studies have used a reflective measurement model (i.e., exploratory or confirmatory factor analysis), wherein changes in a latent construct are thought to precede and in some sense, explain variation in observed variables. Others have argued that engagement is best measured using a formative model in which the relationship flows in the opposite direction. In other words, within formative measurement variation in observed indicators precedes, and can in some sense either create or cause a construct. A clear rationale has not been provided for the use of either measurement model. In the current study, I therefore sought to compare a series of reflective and formative measurement models using the Gallup-Purdue Index (GPI; Gallup-Purdue, 2014), an under-examined national instrument that has defined student engagement as three inter-related, albeit distinct, latent constructs: institutional support, institutional attachment, and experiential learning. For the investigation, data were collected from alumni who attended a mid-sized southeastern university and graduated with a bachelor's degree between 1996 and 2005. The current study occurred within three stages. First, an exploratory factor analysis of GPI engagement items was investigated using a random subsample of 349 respondents. This was followed by the second stage wherein three competing models were tested using confirmatory factor analysis on a random subsample of 700 students. Finally, three formative models were also examined using the second subsample. Results of the analyses provided support for a reflective model of the GPI

engagement items. Implications are offered regarding the use of formative and reflective approaches and the conceptualization of student engagement.

## CHAPTER 1

### **Introduction**

#### **Student Engagement as a Measure of Quality in Higher Education**

The quality of undergraduate education in the United States has undergone increased scrutiny with heightened skepticism, dissatisfaction with the current status of higher education, and calls for reform (Arum & Roksa, 2011; National Commission on the Future of Higher Education, 2006). A range of factors have stimulated interest in the assessment of quality among stakeholders of higher education (Coates, 2005). Students need accurate information about educational quality to make informed decisions about which institutions to attend and desirable courses of study. Faculty and university administrators need information to help them evaluate and improve educational programs. Institutions need information about quality to help them benchmark and market their performance. Legislators need information to assist with funding, policy development, and accountability.

In response to such concerns, promoting student engagement through educationally purposeful activities and experiences has been advocated as an effective way to transform undergraduate education and assess institutional quality (NSSE, 2005). Empirical evidence points to the promise of strategies focused on student engagement. For example, educationally purposeful experiences in college are critical to student learning and personal development (Astin, 1993; Pascarella & Terenzini, 1991, 2005; Paulsen, 2013). Such activities have been connected to the concept of student engagement, which is historically rooted in Astin's (1984) theory of involvement. Astin's theory stated that the amount of student learning and development is proportional to

students' involvement in educational or extracurricular activities. Similarly, the concept of "quality of effort" put forth by Pace (1980, 1982) and Pascarella (1985) has influenced current conceptualizations of student engagement. Quality of effort refers to the claim that the more a student is meaningfully engaged in an academic task, the more the student will learn.

Despite the widespread interest and participation in student engagement related initiatives, there is little consensus about the best way to define and measure the construct. The current research study included two important areas of focus: 1) gathering validity evidence for engagement-related items on the Gallup-Purdue Index (Gallup, 2014), an under-examined national instrument and 2) evaluating how the engagement-related items are best modeled. In this introduction, I first introduce Benson's (1998) validation framework. This is followed by an overview of how student engagement has been defined within the literature and an evaluation of previous psychometric evidence of instruments that have aimed to measure engagement. My evaluation of previous psychometric research leads to an examination of two competing models that could be used to both conceptualize and measure engagement. I conclude the section by describing the purpose of the current study.

### **Benson's Validity Framework**

Validation is one of the most important aspects of instrument development, because it involves the central question of whether the inferences we make from scores on an instrument are useful and appropriate. That is, validity has been defined as, "...the degree to which all the accumulated evidence supports the intended interpretation of tests scores for the proposed use" (AERA, APA, & NCME, 2014). Thus, one study does not

validate or fail to validate the scores from an instrument. Multiple studies may be required, using different approaches and different samples to build a body evidence that supports (or fails to support) the validity of the scores derived from an instrument (Benson, 1998).

Benson's (1998) three-stages of construct validation are a framework for building a body of validity evidence for score use and interpretation. The three stages include: 1) a substantive stage that is concerned with a clear definition of the theoretical and empirical domains of student engagement; 2) a structural stage focused on the dimensionality of the Gallup-Purdue Index student experiences and attachment subscale (GPI-SEAS); and 3) an external stage that emphasized the relationship between student engagement and other constructs, using structural equation modeling.

### **What is Student Engagement?**

A plethora of definitions of student engagement can be found within the literature. The number and range of definitions has led to conceptual ambiguity and confusion. Nevertheless, these definitions can be synthesized to: students' involvement in academic and co-curricular experiences provided by the institution consisting of affective, cognitive, and behavioral investment (Butler, 2011; Chapman, 2003; Handelsman, Briggs, Sullivan, & Trowler, 2005; Kuh, 2003; Mandernach, 2015). A brief overview of how student engagement has been defined is provided below.

Some definitions have drawn connections between engagement and participation whereas other researchers have viewed engagement as a multi-faceted construct. Student engagement has been described as "participation in educationally effective practices, both inside and outside the classroom, which leads to a range of measurable outcomes (Harper

& Quaye, 2009, p. 2). Akey (2006) also stated student engagement is "... the level of participation and intrinsic interest that a student shows" and which involves behaviors, attitudes, and affect (p. 6). Although participation is a critical component of many student engagement definitions, some researchers claimed student engagement is a multifaceted construct. That is, some definitions emphasized that engagement relies not only on choices made by students, but also on the opportunities made available to them by the institution (Harper & Quaye, 2009; Kuh, 2003).

Disagreements concerning the definition of student engagement are typically due to subtle differentiations between engagement as a *process* versus a *product* (Mandernach, 2015). Bowen (2005) contended student engagement can be defined in four ways: 1) engagement with the *learning process* (i.e., active learning); 2) engagement with the *object of study* (i.e., experiential learning); 3) engagement with the *context of study* (i.e., multidisciplinary learning); and 4) engagement with the *human condition* (i.e., service learning). Bowen (2005) claimed most assessments of student engagement emphasize the learning process. However, Barkley (2010) emphasized that "student engagement is the product of motivation *and* active learning. It is a product rather than a sum because it will not occur if either element is missing" (p. 6). Although it may seem subtle, the distinction between student engagement as a process or a product has important implications for the assessment and measurement of the construct. Assessments of engagement as a process should emphasize behaviors, activities, and attitudes that contribute to student learning. Conversely, assessments of engagement as a product emphasize the cognitive or affective state resulting from a learning process (Mandernach, 2015).

In addition to the disagreements about engagement as a process or a product, many researchers have claimed student engagement has three interrelated aspects: cognitive, behavioral, and affective (Butler, 2011; Chapman, 2003; Handelsman et al., 2005; Kuh, 2003; Mandernach, 2015). The cognitive aspect of engagement includes investment in learning and intellectual energy. The behavioral aspect includes involvement in the task at hand, participation, and interactions with others. The affective aspect includes responses (such as interest or anxiety) to instructors, the learning environment, and the institution (Baron & Corbin, 2012). A deeper exploration of student engagement definitions follows in chapter two.

### **Prior Psychometric Evaluations of Student Engagement**

Student engagement is not only difficult to define but also challenging to measure (Ryan, 2005). An effective assessment of student engagement first requires that a researcher or practitioner knows what aspects of student engagement are the focus of inquiry. Once a researcher knows what aspects of student engagement are being targeted, he or she should then investigate the psychometric properties of scores from the measure to bolster the validity of scoring procedures and any inferences made from students' responses to the items.

The most commonly used methodology for student engagement assessment is that of self-report questionnaire (Baron & Corbin, 2012). Instruments such as the College Student Experiences Questionnaire (CSEQ), the National Survey of Student Engagement (NSSE), and the Community College Survey of Student Engagement (CCSSE) have been widely used at colleges and universities throughout the United States over the past few decades. Despite their wide-scale use, a quick review of previous research reveals



psychometric inconsistencies. For example, NSSE researchers initially used a combination of principal components analysis (a data reduction technique) and theory to derive five benchmarks that represented the student engagement items (Kuh, 2009). However, several subsequent studies used factor analysis or a combination of factor analysis and principal components analysis to replicate NSSE's original findings (e.g., Campbell & Cabrera, 2011; LaNasa, Cabrera, & Trangsrud, 2009; LaNasa, Olson, & Alleman, 2007; Swerdzewski, Miller, & Mitchell, 2007). The subsequent studies employed techniques that have different underlying assumptions about the best way to model student engagement. For example, later studies found different solutions (e.g., eight or nine dimensions of student engagement) that did not replicate the original study's five benchmarks (LaNasa, Cabrera, & Trangsrud, 2009; LaNasa, Olson, & Alleman, 2007; Marti, 2009). The different underlying assumptions may in part explain the inconsistent results about the NSSE's dimensionality.

### **Is Student Engagement a Formative or a Reflective Latent Variable?**

Inconsistencies in the measurement properties of scores from the NSSE reinforce the need for all researchers to justify, both theoretically and empirically, their choice of measurement model. Use of an incorrect measurement model can undermine the interpretation of content validity evidence, misrepresent the structural relationships between constructs, and detract from the usefulness of student engagement theory for researchers and practitioners (Coltman, Devinney, Midgley, & Venaik, 2008). Specifically, the use of principal components analysis (PCA) in NSSE analyses leads to different inferences than the use of exploratory factor analysis (EFA). Though the analytic procedures for PCA and EFA are similar (with the exception of the diagonal of

the variance/covariance matrix), they differ in important ways conceptually (Hathcoat & Meixner, 2015). Principal components analysis is a formative measurement model, whereas EFA is a reflective measurement model.

With respect to reflective measurement models, researchers treat indicators as outcomes of a latent variable (Figure 1a). Indicators are observed variables whereas latent variables are defined as hypothetical constructs or explanatory variables that represent a continuum that is not directly observable (Kline, 2011). Referring back to the example of intelligence, there is no single, conclusive measure of intelligence. Rather, researchers use different observed variables such as memory capacity or verbal reasoning to assess areas of intelligence. According to a reflective measurement model, changes in a latent construct precede and in some sense, explain changes in the observed variables. For example, imagine that a group of students were asked to take a test that asked them to identify the logical or mathematical relationship in a series of objects or numbers. Performance on these items are expected to be correlated, thus some students will have high scores whereas other students will have low scores. The researcher hypothesizes that the reason for this pattern is due to differences in intelligence. That is, a person's unobservable (latent) level of intelligence is expected to influence their scores (indicators) on an intelligence test. This is done by statistically relating covariation between the latent constructs and the observed variables or indicators of the latent constructs (Borsboom, Mellenbergh, & van Heerden, 2003).

A reflective measurement model implies that if variation in an indicator is associated with variation in a latent construct, then interventions that change the latent construct can be detected in the indicator (Coltman et al., 2003). In other words,

interventions on intelligence should result in changes in the observed variables. Many studies related to the dimensionality of student engagement measures (e.g., CSEQ, CCSSE, and NSSE) use reflective measurement models, but a clear rationale for the use of these models is rarely provided.

Although the reflective view seems to dominate psychology and education literature, the formative view is more common in sociology and economics (Coltman et al., 2008). However, a few researchers in the psychology literature indicate that not all latent constructs are best measured by a set of positively correlated items (Bollen & Lennox, 1991; Edwards & Bagozzi, 2000). Formative measurement models differ from reflective models in several respects. First, in formative models indicators are combined to form a construct without making assumptions about the pattern of intercorrelations between the items (Figure 1b). In addition, the relationship flows in the opposite direction as the reflective model, from the indicator to the construct (Blalock, 1968; Diamantopoulos & Winklhofer, 2001; Edwards & Bagozzi, 2000). An example of formative measurement is socioeconomic status (SES) measured by such observed variables as education, income, and occupational prestige (Heise, 1972). With respect to SES, it is inappropriate to conceive of SES as a cause of the observed variables, such as occupational prestige. Rather, SES is better viewed as a function of occupational prestige. Viewing SES as a function of occupational prestige implies that changes in occupational prestige precede, and in some sense account for, changes in SES. Interventions in a formative model would therefore focus on observed variables as opposed to the latent variable.

The distinction between reflective and formative measures is important because proper specification of a measurement model is necessary to assign meaningful relationships to outcomes or other constructs (i.e., via a structural model; Anderson & Gerbing, 1988). For example, model misspecification can result in biased parameter estimates and interpretational confounding (see chapter two for a more detailed explanation; Bollen, 1989, 2007). Thus, considering both formative and reflective approaches for the student engagement construct may help us avoid potential issues of model misspecification.

### **Purpose of the Current Study**

In response to questions about the quality of higher education, Gallup, Inc. partnered with Purdue University and the Lumina foundation to conduct a national study evaluating the long-term success of college graduates. The resulting instrument, the Gallup-Purdue Index (GPI) links alumni's perceptions of their college experiences to their current well-being, life satisfaction, and careers (Gallup, 2014). One of the GPI subscales (i.e., GPI-SEAS) asks alumni questions related to their participation in educationally purposeful activities and experiences while they were in college as well as their perceptions of the institution. The GPI was first administered in 2014 and many of the instrument's subscales, including the GPI-SEAS, do not have published information about their psychometric properties. The GPI-SEAS subscale is unique to other student engagement measures (such as the CSEQ or the NSSE) in that its methodology provides the opportunity to examine engagement in relation to post-college outcomes such as life and work satisfaction without using a longitudinal sample.

Although many researchers have taken a reflective approach to modeling student engagement, little justification has been provided for why that strategy is appropriate. The current study includes an examination of validity evidence for the GPI-SEAS items and an investigation of the structure of the GPI-SEAS items and their relation to post-college outcomes. Currently, there is no published research related to the dimensionality and validity evidence for the GPI-SEAS. Specifically, in the current study, I investigated the appropriateness of either a reflective or formative measurement model for the GPI-SEAS as a measure of student engagement. As detailed in chapter two, the research questions for the current study related to the internal structure of the GPI-SEAS and its relation to outcome variables, such as workplace satisfaction and general life satisfaction. The current study contains a discussion of the theoretical history of student engagement in order to define the construct of student engagement. That is, it is argued that aspects of student engagement may be better conceived as a formative construct than as a reflective construct. The introduction to engagement and its measurement is followed by a series of empirical tests to investigate the directional relationship between the construct and its observed indicators.

## CHAPTER 2

### Literature Review

Researchers have studied student engagement in higher education for decades. However, it remains unclear how student engagement should be conceptualized and modeled psychometrically. In this literature review, I argue that aspects of engagement may be better conceived as a formative measurement model rather than a reflective measurement model. This literature review also includes background information about the construct of student engagement. Specifically, the review includes an investigation of various theories of student engagement, definitions of student engagement as described in the higher education literature, an examination of measures of student engagement, and the relationship of student engagement to other constructs. Information on current definitions of student engagement, the measurement of student engagement, and their relationship to other constructs is crucial to the empirical definition of student engagement used in the current study. In addition, the review includes a description of the Gallup-Purdue Index, including its background, purpose, and details of the subscales used for the current study. Finally, the review includes an explanation of relevant data analytic techniques, including discussions of formative and reflective measurement along with challenges associated with each approach. Within this section, I also discuss criteria for choosing between formative and reflective approaches, using student engagement as an example, to determine how the construct should be modeled in the current study.

### Engagement Theory and Research

**Defining student engagement.** The term *engagement* has permeated higher education literature since the mid-1990s, with civic engagement, the scholarship of

engagement, and student engagement included in the discourse. However, these various uses of the term mean different things. For instance, civic engagement refers to the ways in which colleges and universities focus on students' proclivity towards advancing the well-being of their local communities through political and non-political means (Bingle, Games, & Malloy, 1999; McCormick, Kinzie, & Gonyea, 2013; Saltmarsh & Hartley, 2011). Student engagement has been defined differently by researchers over the past few decades. Most definitions of student engagement include students' participation in learning activities as a critical component of engagement. Yet, many authors have described engagement as a multidimensional phenomenon.

Current definitions of student engagement include cognitive, affective, and behavioral components of student engagement (Butler, 2011; Chapman, 2003; Handelsman et al., 2005; Kuh, 2003; Mandernach, 2015). That is, student engagement is a term that can include the extent to which students participate in educationally effective activities. Student engagement can also include students' perceptions of aspects of the institutional environment that support their learning and development (Kuh, 2009; McCormick, Kinzie, & Gonyea, 2013; Harper & Quaye, 2015). For example, Astin (1984) emphasized that the cognitive aspect of engagement involves not only a behavioral investment of time, but also requires the investment of attention and intellectual energy. Skinner and Belmont (1993) underscored the behavioral and affective aspects of learning and defined student engagement as "sustained behavioral involvement in learning activities accompanied by positive emotional tone" (p. 572). Although they focused on engagement at a K-12 school level, Fredricks, Blumenfeld, and Paris (2004), drawing on Bloom (1956), identified three dimensions of student engagement: behavioral

engagement, emotional engagement, and cognitive engagement. Behavioral engagement refers to students who participate in academic, social and/or extracurricular activities. Emotional engagement refers to the experience of affective reactions within activities such as interest, enjoyment, and a sense of belonging. Cognitive engagement refers to an investment in learning and seeking to go beyond minimal the minimal requirements of a course (Fredricks et al., 2004; Trowler, 2010).

Similarly, McCormick et al. (2013) contended that student engagement incorporates behavioral and perceptual components. They described the behavioral dimension as including how students use their time in and outside of the classroom (e.g., collaborating with peers in learning activities, interacting with faculty, and integrating ideas across courses). Because attitudes and beliefs are antecedents to behavior (Bean & Eaton, 2000), students' perceptions of the campus environment are an important aspect of assessing students' openness to learning. The perceptual dimension of student engagement includes students' judgments about their relationships with peers, faculty, and staff; their beliefs regarding faculty expectations of students; and their understanding of institutional norms concerning academic activities and support of student success (McCormick et al., 2013).

In addition to the cognitive, affective, and behavioral aspects of engagement, Harper and Quaye (2009) argued that student engagement hinges on both the institution and the student. That is, students must make the choices to participate in educational activities, but faculty and student affairs professionals must make the opportunities available through the institution. Kuh (2003) provided a definition of student engagement that integrates the affective, cognitive, and behavioral components while also



highlighting the dual role of students and the institution to foster engagement. According to Kuh (2003), student engagement is “the time and energy students devote to educationally sound activities inside and outside the classroom, and the policies and practices that institutions use to induce students to take part in these activities” (p. 25).

Based on a unification of definitions from previous research, this study focuses on student engagement as students’ involvement in academic and co-curricular experiences provided by the institution. The academic and co-curricular experiences consist of affective, cognitive, and/or behavioral investment (Butler, 2011; Chapman, 2003; Handelsman et al., 2005; Kuh, 2003).

**Theoretical conceptualizations of student engagement.** Conceptualizations of student engagement hinge on impactful educational practices: the experiences and activities empirically linked to desired college outcomes. Historical influences of student engagement go back to the 1930s and include areas of sociology, psychology, cognitive development, learning theory, and higher education research. The meaning of the construct has also evolved over time. One of the earliest iterations can be traced to educational psychologist Ralph Tyler in the 1930s, who showed the positive effects of time on task to learning (Merwin, 1969). Tyler’s work was explored more thoroughly by Pace (1980, 1990), who showed that the time and energy students invest in educationally-relevant tasks (e.g., studying, interacting with peers and faculty, and applying what they learn to tasks outside of class) is a key factor in student success. Pace developed the College Student Experiences Questionnaire (CSEQ), which was based on what he termed “quality of effort” – the claim that the more a student is meaningfully engaged in an academic task, the more the student will learn.

Alexander Astin (1984) furthered the quality of effort concept with his developmental theory of involvement. He defined quality of effort as “the amount of physical and psychological energy that the student devotes to the academic experience” (p. 297). Astin also contended that the amount of student learning and development is proportional to students’ involvement in the educational or extracurricular program. He recognized that involvement may be similar to the concept of motivation, but differs in that motivation is a psychological state, while involvement indicates behavior. The ideas of time spent on task and quality of effort put forth by Tyler, Pace, and Astin, all contribute to current conceptualizations of student engagement.

Both Pace (1980) and Astin (1984) emphasized the important role of the institutional environment and what the college or university contributes (or fails to contribute) to enhancing student effort and involvement. Pace (1982) thought of students as active participants in their own learning who must take advantage of the educational resources provided by their campus. Although students have a responsibility in using the resources and participating in opportunities available to them, Astin (1984) highlighted the critical role of the institution. That is, he asserted, “the effectiveness of any educational policy or practice is directly related to the capacity of that policy or practice to increase student involvement” (p. 298).

Another major contribution to the conceptualization of student engagement is Tinto’s theory of academic and social integration (1975, 1993). The term *integration* refers to the extent to which a student (a) comes to share the attitudes and beliefs of peers and faculty and (b) follows the rules and requirements of the institution (Pascarella & Terenzini, 1991; Tinto, 1975, 1993). Social integration refers to students’ perceptions of

their interactions with peers, faculty, and staff as well as their involvement in extracurricular activities. Academic integration refers to students' academic performance, identification with academic norms, and compliance with standards of the college or university. Tinto was one of the first researchers to theorize that voluntary student departure involved not just the student, but also the institution. Influenced by Tinto's work, current conceptualizations of student engagement include students' interactions and connections with peers and faculty as well as the extent to which they use academic resources and feel supported at the institution.

Pascarella (1985) expanded on Tinto's research by including institutional characteristics and the quality of student effort. Pascarella also linked his research to outcomes other than retention, which pervaded the higher education literature. He posited that students' precollege characteristics correlated with type of institution. Further, precollege characteristics and type of institution were related to the institutional environment and students' meaningful interactions with peers, faculty, and institutional administrators. Pascarella maintained that student background and precollege characteristics, institutional environment, and interactions with peers, faculty, and staff influenced quality of effort. Both Tinto's and Pascarella's focus on students' interactions with their institutions and institutional values and norms provide the foundation for the environmental aspects of the student engagement concept.

Teaching and learning literature has also contributed to conceptualizations of student engagement. Chickering and Gamson (1987) synthesized 50 years of higher education research into seven principles for teaching and learning. These principles included, (1) contact between faculty and students, (2) reciprocity and cooperation among

students, (3) active learning, (4) providing prompt feedback, (5) emphasizing time on task, (6) communicating high expectations to students, and (7) respecting diverse talents and ways of learning (Chickering & Gamson, 1987). They contend that each principle can stand on its own, but in combination, the effects multiply and can have a powerful impact on students' educational experiences. Chickering and Gamson (1987) also emphasized the responsibility of educators and university leaders to foster an environment that supports good practice and to ensure that students regularly engage in effective educational practices. Longitudinal analyses at a diverse group of 18 institutions supported the relationship between the principles of teaching and learning and cognitive development and other positive outcomes. The findings of the longitudinal analyses suggest that environmental conditions at an institution may facilitate student engagement (Cruce, Wolniak, Seifert, & Pascarella, 2006).

**Recent developments in student engagement research.** In more recent years, activities deemed “high-impact practices” such as service learning, internships, learning communities, and undergraduate research, have been identified as indicators of student engagement (AAC&U, 2007; Kuh, 2008a). High-impact practices (HIPs) require that students devote considerable time and effort to purposeful tasks and often require close interaction with faculty and diverse students (Kuh, 2008b). HIPs, such as study abroad, internships, capstone experiences, and collaborative projects, invite students to apply their learning in innovative ways through problem solving with peers inside and outside the classroom. For example, Zhao and Kuh (2004) showed that students who participated in learning communities were more engaged in other educationally purposeful activities than students who did not participate in learning communities. Learning community

students interacted more with faculty and diverse peers, studied more, and reported gaining more from their college experience than other students. In addition, Rocconi's (2011) study showed through path analytic techniques, that learning community participation was indirectly related to educational gains through student engagement. That is, interactions with faculty members, effort in coursework, and interactions with peers were positively related to students' educational gains.

Within the last decade, scholars have also contributed to understandings of student engagement from an instructional perspective. For instance, Gabriel (2008) explained the value of student engagement for teaching underprepared students. Other educators and researchers (e.g., Ahlfeldt, Mehta, & Sellnow, 2005; Barkley, 2010; Smith, Sheppard, Johnson, & Johnson, 2005) investigated classroom-based models of engagement, particularly problem-based and active learning that focus on student involvement in the learning process. These examples emphasize the connection between student engagement and educational practice and highlight the commitment of educators to improvement guided by classroom-based evidence.

Throughout the past 50 years, higher education research on academic and social integration, involvement, quality of effort, and best practices in education, indicates conditions that support student engagement require the commitment of students, individual faculty members, and the institution as a whole. Students must:

- commit to putting forth quality effort,
- get involved in educational experiences inside and outside of the classroom,
- make decisions about how to best allocate their effort in coursework, and
- interact with other students informally or formally through co-curricular activities.

Faculty must:

- commit to providing learning opportunities and activities in their courses
- clearly convey their expectations to students
- provide useful feedback to students, and
- facilitate student learning outside of the classroom through formal or informal means.

Institutional staff and administrators must:

- create standards that support student success and
- allocate resources to support student success.

As an example, library and student affairs professionals have created supportive learning environments through programs and events that enrich undergraduate students' experiences (Gilchrist & Oakleaf, 2012; Quaye & Harper, 2014; Strange & Banning, 2001). Some institutional leaders have also established policies and practices that communicate standards for students, faculty, and staff pertaining to student support (Donald, 1997; Grunwald & Peterson, 2003). The study of student engagement not only promotes student success, but also contributes to larger national conversations about college impact and quality.

### **Measuring Student Engagement**

**Student engagement instruments.** The aforementioned research and literature provided definitions of student engagement that were related, but in diverse ways. Many of the instruments designed to measure student engagement focus on the behavioral and perceptual/emotional aspects of engagement. A few of the most popular and widely-used instruments include: The College Student Experiences Questionnaire, the College Student

Expectations Questionnaire, and the suite of NSSE instruments, which include the National Survey of Student Engagement, the Beginning College Survey of Student Engagement, the Community College Survey of Student Engagement, and the Faculty Survey of Student Engagement. Finally, the Gallup-Purdue Index, a recently-developed instrument on alumni perceptions of their college experiences, perceptions of college worth, and current well-being, is described given that this instrument is the primary measure of interest in the current study.

*College Student Experiences Questionnaire.* The College Student Experiences Questionnaire (CSEQ) was developed by C. Robert Pace at the University of California, Los Angeles (UCLA) and introduced in 1979. Pace (1980, 1982) believed measuring students' quality of effort would help researchers and educators better understand student learning and development. The CSEQ was designed to measure the "quality of effort students expend in using institutional resources and opportunities provided for their learning and development" (CSEQ, 2007, para. 1). The goal of the CSEQ is to assess students' perceptions of the overall learning and institutional environment to provide formative and diagnostic feedback to faculty and administrators (Gonyea, Kish, Kuh, Muthiah, & Thomas, 2003; Williams, 2007).

The most recent version of the CSEQ includes three sections related to student engagement – college activities, opinions about your college or university, and the college environment. *College activities* consists of 13 subscales (109 total items) requiring students to report on the quality of effort they expend on activities related to:

- library experiences (8 items);
- computer information and technology (9 items);

- course learning (11 items);
- writing experiences (7 items);
- experiences with faculty (10 items);
- art, music, and theater (7 items);
- campus facilities (8 items);
- clubs and organizations (5 items);
- personal experiences (8 items);
- student acquaintances (10 items);
- scientific and quantitative experiences (10 items);
- topics of conversation (10 items; e.g., discussed current events, social issues, or the arts outside of class); and
- information in conversations (6 items; e.g., whether the student discussed class readings, explored different ways of thinking, or referred to something their instructor said outside of class).

The *Opinions about Your College or University* section includes two items, and there are 10 items about the *College Environment* (i.e. questions about students' perceptions of the extent to which the campus emphasizes diverse learning experiences and relationships with faculty, administrators, and other students; Williams, 2007). Internal consistency, as indicated by coefficient alpha ranged from .70 to .92 across each of the subscales in the *College Activities* section. Alpha values of .70 and above indicate adequate reliability for applied research (Gonyea et al., 2003). Psychometric evaluations of scores from the measure included principal axis factor analysis of the 13 college activities subscales, each of which resulted in one-factor solutions (Gonyea et al., 2003). In other words, a factor



analysis was conducted separately for each subscale. It is unclear why factor analyses were conducted on each individual subscale rather than the entire set of college activities questions to determine if evidence supported the 13-factor structure. Details of how the models were estimated were not included in the study, so results should be interpreted with caution.

When Pace retired from UCLA in 1993, the CSEQ was transferred to Indiana University Bloomington under the direction of George Kuh. The CSEQ was a popular and widely used measure in higher education for 35 years. Since its initial administration in 1979, over 500 institutions and researchers used the CSEQ and over 400,000 students completed the questionnaire. The College Student Expectations Questionnaire (CSXQ) was launched in 1998 as an extension of the CSEQ. The CSXQ was designed to assess first-year student goals and motivations (CSXQ, 2007). The measure was developed by C. Robert Pace and George Kuh, who believed first-year students hold important expectations about how and with whom they will spend their time in college (Williams, 2007). Since its initial implementation, the CSXQ was completed by 120,000 students from over 100 institutions (CSXQ, 2007). The CSEQ and CSXQ survey operations were closed in 2014 after 35 years of continuous administrations. The widespread use of another measure, the National Survey of Student Engagement (developed by George Kuh), led to the eventual closure of the CSEQ and CSXQ operations.

*National Survey of Student Engagement.* In 1998, Russ Edgerton (1997) of the Pew Charitable Trusts proposed a grant project to improve higher education. Edgerton organized a group of scholars to explore the extent to which colleges and universities emphasize effective teaching practices and students engage in educationally purposeful

activities. The National Center for Higher Education Management Systems (NCHEMS) coordinated the design of the National Survey of Student Engagement (NSSE, pronounced “Nessie”) with support from Pew Trusts. The survey design team included Alexander Astin, Gary Barnes, Arthur Chickering, Peter Ewell, John Gardner, George Kuh, Richard Light, Ted Marchese, and C. Robert Pace (NSSE, 2001). The NSSE design team included many of the most influential student engagement researchers from throughout the previous twenty years. Although operations of Pace’s CSEQ closed, about two-thirds of the original NSSE questions were drawn or adapted from the CSEQ (CSEQ, 2007).

The NSSE was designed to query undergraduate students directly about their educational experiences. The survey is administered during the spring semester as either a sample or census of first-year and senior students. The NSSE survey includes both behavioral and affective components. Items were selected based on their relationship to student learning and development (Ewell, 2010; McCormick et al., 2013). The survey includes 42 key questions grouped into 5 benchmarks: level of academic challenge, active and collaborative learning, student-faculty interaction, enriching education, and supportive campus environment. The *Level of Academic Challenge* benchmark consists of 11 questions that focus on academic effort such as students’ time spent preparing for class (e.g., studying, reading, and writing) and whether students worked harder than they thought they could to meet an instructor’s expectations. The *Active and Collaborative Learning* benchmark consists of seven questions related to working with others to solve problems or master material. Examples of active and collaborative learning include asking questions in class and contributing to class discussions. The *Student-Faculty*

*Interaction* benchmark consists of six items related to students' interactions with faculty inside and outside the classroom. Sample items include, whether students have discussed career plans with a faculty member and whether they have worked with a faculty member on a research project. The *Enriching Education* benchmark consists of 12 items about students' experiences with diversity and participation in activities such as internships, community service, or study abroad. Finally, the *Supportive Campus Environment* benchmark consists of six items concerning students' perceptions of the campus environment and their social relations with different groups on campus. Supportive campus environment items ask students about the quality of their relationships with other students, faculty members, and staff, as well as whether the campus environment provides the support necessary for them to succeed academically.

Specific classroom activities, as well as faculty and peer practices, are positively related to student outcomes (Ahlfeldt et al., 2005; Barkley, 2010; Smith et al., 2005). The degree to which students are engaged in their studies is directly related to the quality of student learning and the overall educational experience (NSSE, 2001). Given such evidence, the NSSE team contends that characteristics of student engagement can serve as a proxy for quality (NSSE, 2001). That is, the NSSE provides an alternative to college rankings by collecting information from students that has the potential to reframe the local and national conversations about institution quality. The NSSE developers proposed three possible uses for the data. First, results should be useful to the institutions collecting the data to improve undergraduate education. Second, aggregated results should be beneficial to external stakeholders of higher education such as accreditors and state

education agencies. Third, if survey results are made public, they might be appealing to the news media and creators of college guides (NSSE, 2001).

NSSE has become the most widely used student engagement measure in higher education and approximately 5.5 million students completed the survey between 2000 and 2016 (NSSE, 2016). NSSE is now a self-supporting auxiliary unit within the Center for Postsecondary Research (CPR) in the Indiana University School of Education. As a result of NSSE's popularity and perceived usefulness in higher education, other CPR surveys have been developed as extensions of NSSE, including the Beginning College Survey of Student Engagement, the Community College Survey of Student Engagement, and the Faculty Survey of Student Engagement. Each measure is described briefly in the subsequent sections.

*Psychometric Evaluations of Scores from the NSSE.* The psychometric properties of scores from the NSSE have been extensively evaluated, but have produced inconsistent results. The five benchmarks were constructed with “a blend of theory and empirical analysis” (Kuh, 2009). NSSE researchers initially used principal components analysis (PCA) with oblique rotation on scores from a national sample of student respondents (Kuh et al., 2001). Theory was then used in conjunction with the results to assign each of the 42 items to one of five components. Although the benchmarks were constructed at least in part using PCA, the NSSE literature (Kuh, 2009) consistently refers to the benchmarks as “factors.” This implies that the benchmarks represent latent traits rather than a linear combination of observed variables, such as what would be accomplished with data reduction procedures such as PCA (Swerdzewski, Miller, & Mitchell, 2007).

A few researchers have investigated the dimensionality of the NSSE in order to provide validity evidence for the five benchmarks. Porter (2011) reviewed the literature examining the reliability and validity of the NSSE benchmarks using Kane's (1992, 2001) argument-based approach to validity.<sup>1</sup> Porter concluded the NSSE did not meet reliability or validity standards. Specifically, Porter asserted that the structure of the five dimensions of engagement represented by the NSSE benchmarks had not been replicated and reliability values failed to meet basic standards (i.e., coefficient alpha levels at or above .70). For example, two groups of researchers (LaNasa, Cabrera, & Trangsrud, 2009; LaNasa, Olson, & Alleman, 2007) used factor analysis and found eight separate dimensions of student engagement at one institution. In addition, Swerdzewski et al. (2007) used confirmatory factor analysis and found a five-factor solution reflecting the benchmarks produced poor model fit.

Similarly, Campbell and Cabrera (2011) used confirmatory factor analysis on data from a sample of 1,026 students to investigate the validity of the five NSSE benchmarks. They found that the five-benchmark model did not fit their sample of students and resulted in high intercorrelations among the benchmarks (e.g., two intercorrelations greater than .70), low item loadings (e.g., eight items with loadings less than .30), and low reliability values (i.e., three below .70). In response to this evidence, McCormick (current director of the NSSE) and McClenney (2012) claimed that it was inappropriate to treat the benchmarks as the type of latent construct represented by exploratory and confirmatory factor analytic procedures. That is, McCormick and McClenney claimed the

---

<sup>1</sup> Kane's approach is firmly grounded in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014), a manual issued jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education.

benchmarks instead reflect “summative indices of a range of education practices” (McCormick & McClenney, 2012, p. 324). The five NSSE benchmarks were created from the NSSE survey items using a combination of theory (specific practices that seem to have the most impact on student outcomes) and PCA. McCormick and McClenney (2012) contended that the benchmarks represent a blend of empirical analysis and expert judgment, rather than latent constructs modeled by factor analysis.

Although a rationale for the benchmarks has been provided, it does not seem appropriate or justifiable. That is, the initial creators of the NSSE used PCA to examine the dimensionality of the instrument (i.e., Kuh, 2009) and subsequent researchers have failed to replicate the five factor solution (e.g., Campbell & Cabrera, 2011, LaNasa, et al., 2007; LaNasa et al., 2009; Porter, 2011, Swerdzewski et al., 2007). McCormick’s arguments against viewing the benchmarks as reflective latent constructs might hold more weight if he provided a stronger rationale for the development of the benchmarks or if he proposed another type of analysis to investigate the hypothesized benchmarks. If the benchmarks are the most effective way to represent student engagement according to the research, and they are not reflective latent constructs, perhaps other types of analyses should be employed to investigate their validity. If other researchers treated the benchmarks as formative latent constructs, results of additional analyses might provide support for McCormick and McClenney’s (2012) argument.

*Beginning College Survey of Student Engagement.* The Beginning College Survey of Student Engagement (BCSSE, pronounced “Bessie”) is similar to the CSXQ in that it assesses engagement dimensions of students entering college. The BCSSE was launched in 2007 and revised in 2013 to align with an updated version of the NSSE (Cole & Dong,

2013). The BCSSE examines the expectations of beginning college students for participating in academic activities and initiatives via nine scales. Cole and Dong (2013) used BCSSE data collected from over 70,000 students to conduct confirmatory factor analysis on scores from the nine subscales. A nine-factor model adequately fit the data. Two of the scales referred to students' academic engagement in high school quantitative reasoning and learning strategies. Three of the scales included students' first-year expectations to engage in collaborative learning with peers, interactions with faculty, and interactions with diverse students. The other four scales addressed students' expected academic perseverance, expected academic difficulty, perceived academic preparation, and the importance of the campus environment to support their academic efforts (Cole & Dong, 2013). BCSSE administration should occur prior to the start of fall classes for first year students and is designed to be paired with administration of the NSSE in the spring semester. The data from the two surveys can provide indicators of the extent to which institutions have met students' expectations regarding engagement (Mandernach, 2015).

*Community College Survey of Student Engagement.* The Community College Survey of Student Engagement (CCSSE) was adapted from the NSSE in 2001 with support from the Pew Trust and the Lumina Foundation (McCormick et al., 2013). The CCSSE was designed to examine the unique missions, objectives, and student populations of two-year community colleges (Manderbach, 2015; McClenney, Marti, & Adkins, 2006). Like NSSE, CCSSE is administered in the spring semester, but irrespective of a student's year in school. Instead of academic year, CCSSE collects information about the number of credit hours earned by each student (McCormick et al., 2013). Confirmatory factor analyses of scores from the survey were conducted and

resulted in a nine-factor structure (Marti, 2009). Marti stated that confirmatory analyses were used after exploratory analyses. However, no information was included about the exploratory analyses. Specific information about how the confirmatory factor models were estimated was not included either. Therefore, the results should be interpreted with caution.

Marti (2009) described a similar process to that of the NSSE that led to the CCSSE benchmarks:

Constructing a latent variable model with the best fit to the data and creating latent constructs useful for evaluating the engagement of a student body are clearly complementary efforts. Nevertheless, the two goals diverge, as optimal model fit requires a granular model of latent constructs whereas establishing benchmark measures is a molar endeavor that seeks to broadly classify items with less concern for the precision of model fit. (p. 5)

The CCSSE Technical Advisory Panel reviewed the CFA results and assigned items to benchmarks, taking into account the factor analysis results, reliability estimates, and expert judgment based on theory and empirical evidence related to undergraduate student engagement and learning. The panel review resulted in a five-benchmark structure for the CCSSE items (Marti, 2009). The five constructs included active and collaborative learning, student effort, academic challenge, student-faculty interaction, and support for learners. It is unclear why the original nine-factor structure deemed adequate through factor analysis was not retained or why a final five factor structure was ultimately deemed appropriate by the panel. Marti (2009) reported the five-factor structure exhibited



adequate model fit. However, details of this subsequent analysis were not included in the report. Again, these results should be interpreted with caution.

### **The Gallup-Purdue Index**

Amidst questions about the quality of a college degree and its impact on the lives of graduates, Gallup, Inc. partnered with Purdue University and the Lumina Foundation to conduct a nationally representative study of college graduates. The Gallup-Purdue Index (GPI) launched in 2014 and was designed to evaluate the long-term success of college graduates as they pursue “great jobs and great lives” (Gallup, 2014, para. 1).

The GPI provides the opportunity to collect information related to student experiences and link them to post-collegiate outcomes, such as life and work satisfaction, workplace engagement, and career earnings. The GPI evaluates college alumni, including items related to employee engagement (the Gallup Employee Engagement Index; Gallup, 2016), well-being (the Gallup-Healthways Well-Being 5 View; Gallup, 2014a), life satisfaction, finances and student loans, extracurricular activities engaged in while in college, work/job satisfaction, and graduates’ experiences during college and their attachment to the institution. The GPI was the first instrument of its kind used to conduct a large-scale, nationally representative study of college graduates and their long-term outcomes.

The subscale of interest (i.e., the Gallup-Purdue Index student experiences and attachment subscale [GPI-SEAS]) in the current study includes questions related to graduates’ experiences during college and their attachment to the institution. This GPI-SEAS includes 10 items grouped into three categories: three items Gallup refers to as “experiential learning” (e.g., “While attending [institution], I worked on a project that

took a semester or longer to complete.”), three items Gallup refers to as “support” (e.g., “My professors at [institution] cared about me as a person.”), and four items Gallup refers to as “attachment to the institution” (e.g., “[institution] was the perfect school for people like me.”). Experiential learning, support, and attachment relate directly to the behavioral and affective domains of student engagement described previously in this chapter. Gallup has avoided labeling this set of questions as *student engagement*. However, the experiential learning, support, and attachment items resemble questions labeled as engagement on Gallup’s K-12 instrument, the Gallup Student Poll (Lopez, Agrawal, & Calderon, 2010), as well as some of the NSSE’s student engagement items. Researchers have studied the measurement properties including dimensionality of a few of the GPI subscales, such as the Gallup Employee Engagement Index (Gallup, 2016) and the Gallup-Healthways Well-Being 5 View (Rath & Harter, 2010, 2011; Sears et al., 2014). However, the experiential learning, support, and attachment questions do not have any published research on the item construction, reliability, or validity.

The GPI provides the possibility to address a few different issues related to student engagement. First, we have the ability to evaluate how student engagement should be measured (i.e., using the experiential learning, support, and attachment items). Second, we have the unique opportunity to connect student engagement to post-college outcomes such as life satisfaction and work satisfaction. Given the growing concern about the value of a college education, it is notable that we have a chance to evaluate experiences during college in relation to life after college (after first, making sure we are measuring the construct appropriately).

## **Student Engagement and Related Outcomes**

Much of the literature on student engagement has focused on its relation to educational outcomes and student development during college. It is well established in higher education literature that student engagement is a fundamental part of a quality education and plays an important role in many desirable outcomes (Astin, 1993; Hu & Kuh, 2003; Hu, Scheuch, Schwartz, Gayles, & Li, 2008; McCormick, et al., 2013; Pascarella and Terenzini, 1991, 2005). In the following section I touch on learning and developmental outcomes and their relationship with student engagement during college. However, the bulk of this section focuses on post-college outcomes linked to student engagement, as those outcomes are the focus of this study. Other higher education research provides a more in-depth discourse of student engagement than is presented here in relation to outcomes during college, such as grades, persistence, and critical thinking. (e.g., Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008; McCormick et al., 2013; Trowler, 2010; Trowler & Trowler, 2010).

**Learning and developmental outcomes.** Substantial research connects student engagement to key college outcomes such as learning and development. Positive relationships have been shown between engagement and outcomes such as academic performance or GPA (Carini, Kuh, & Klein, 2006; Gordon, Ludlum, & Hoey, 2008; Kuh et al., 2008), persistence (DeSousa & Kuh, 1996; Kuh et al., 2008), leadership development (Pascarella et al., 2008; Posner, 2004), identity development (Harper, Carini, Bridges, & Hayek, 2004; Hu & Kuh, 2003), moral development (Pascarella et al., 2008), and critical thinking skills (Anaya, 1996; Pascarella et al., 2008; Kuh et al., 2008). Studies that link student engagement to college outcomes help faculty and institutional

leaders understand student success so they can make necessary changes within an institution. Information related to student outcomes provides evidence to support designing faculty development programs, revising curricula, developing student support programs, and redirecting funds and other resources to areas where they can have an impact on students.

**Post-college outcomes.** Although substantial research has identified the important connection between student engagement and learning outcomes, few studies have considered student engagement and post-graduation outcomes. It is certainly time and resource intensive to conduct studies on post-college outcomes, as they necessitate longitudinal data from college and university alumni. With that said, a few studies have been able to show connections between student engagement and post-college outcomes such as life satisfaction, career earnings, and workplace engagement and satisfaction.

**Life satisfaction.** Student engagement has been positively associated with alumni life satisfaction. For example, the supportive campus environment benchmark of the NSSE was positively related to life satisfaction a few years after students graduated from college (Schmaling & Guy, 2014). The supportive campus environment benchmark is a six-item scale measuring the extent to which students feel their campus helps them succeed academically and socially; assists them in coping with nonacademic responsibilities; and promotes supportive relations among students and their peers, faculty members, and administrative personnel and offices (Kuh, 2009). Schmaling and Guy (2014) surveyed 72 college alumni and used zero-order correlation coefficients to explore longitudinal associations between respondents' NSSE benchmarks as college seniors and measures of satisfaction as alumni one to two years after graduation. There

was a moderate, positive relationship between the supportive campus environment benchmark and items that asked whether respondents considered the institution to be committed to student success as alumni ( $r = .37, p < .01$ ).

***Career earnings.*** Early career earnings of college graduates have also been linked to student engagement. Participating in educationally purposeful academic and social activities is related to career earnings when moderated by students' background characteristics (i.e., gender, race/ethnicity, and alumni's pre-college preparation; Hu & Wolniak, 2010, 2013). Hu and Wolniak (2013) extended their earlier (2010) study and looked at longitudinal surveys collected from 1,278 students who participated in the Gates Millennium Scholars program. Students completed surveys as entering freshmen, three years later, and five years after the initial survey. The analysis included eight student engagement items representing two engagement domains (academic engagement and social engagement). The coefficient alpha reliability estimates for each engagement domain were .75 (academic engagement) and .78 (social engagement), which is considered adequate reliability. Hu and Wolniak (2013) used multiple regression analyses with interaction terms to look at group differences in the relationship between student engagement and career earnings. However, it is necessary to differentiate between academic engagement (e.g., working with peers and faculty outside of class on coursework) and social engagement (e.g., participating in extracurricular and community service events) when explaining career outcomes (Hu & Wolniak, 2013). In sum, academic and social systems of student engagement relate to labor market outcomes differently depending on alumni demographics and pre-college characteristics (Hu & Wolniak, 2013).

*Workplace engagement and satisfaction.* Recent studies by collaborators Gallup, Inc. and Purdue University (Gallup-Purdue, 2014, 2015) showed engagement during college is related to job satisfaction and workplace engagement later on. Gallup-Purdue's 2014 survey of over 30,000 college graduates across the United States, indicated students who felt supported at their institution and who participated in experiential learning, were more engaged and satisfied in their jobs than students who did not have those experiences during college (Ray & Kafka, 2014). For instance,

... if graduates recalled having a professor who cared about them as a person, made them excited about learning, and encouraged them to pursue their dreams, their odds of being engaged at work more than doubled, as did their odds of thriving in all aspects of their well-being. And if graduates had an internship or job in college where they were able to apply what they were learning in the classroom, were actively involved in extracurricular activities and organizations, and worked on projects that took a semester or more to complete, their odds of being engaged at work doubled as well (Ray & Kafka, 2014, para. 4).

A follow-up study in 2015 of over 29,000 college graduates affirmed the 2014 findings (Gallup-Purdue, 2015). That is, indicators of student engagement, particularly support and experiential learning were positively related to job satisfaction and workplace engagement. Specifically, college graduates odds of being engaged at work were two times higher if they strongly agreed to the experiential learning and support items.

### **Problems when Measuring Student Engagement**

Despite wide-scale use of student engagement measures, reviews of previously published studies on the NSSE, CCSSE, and CSEQ reveal psychometric inconsistencies.

Published papers on the NSSE and CCSSE have suggested anywhere from five to nine factors of student engagement (Campbell & Cabrera, 2011; Cole & Dong, 2013; Kuh, 2008; LaNasa, et al., 2007; LaNasa et al., 2009). Meanwhile, the CSEQ consists of 13 subscales (each with a single factor structure) related to quality of effort, which serves as an indicator of student engagement. These conflicting conceptualizations (i.e., as a set of related factors that represent the construct or as a composite with multiple measures) also indicate there may be confusion about how to measure student engagement. Many researchers have used factor analytic approaches to measure student engagement (e.g., Campbell & Cabrera, 2011; Cole & Dong, 2013; Kuh, 2008; LaNasa, et al., 2007; LaNasa et al., 2009). While others have used combinations of factor analysis and principal components analysis (Kuh, 2009). Is student engagement a latent trait represented by different factors or is student engagement better conceived as a composite of measures representing students' behaviors and their involvement in different activities?

To answer this question, it is necessary to consider relevant theory related to student engagement. The process of theory development involves emphasis on the relationships among constructs, such as the direction and form these relationships may take and explains why and under what conditions these relationships occur (Edwards, 2011). The process of theory development also involves identifying relationships between constructs and measures (Edwards & Bagozzi, 2000). These relationships represent supplementary theories that connect abstract theoretical constructs to observable phenomena (Blalock, 1968).

When supplementary theories are developed, one of the most fundamental considerations involves specifying the direction of the relationship between constructs and measures (Edwards, 2011). One option is to treat constructs as causes of observed variables, such that the observations are reflective indicators of underlying constructs. This is referred to as a reflective measurement model. Reflective measurement treats observed indicators, such as items, as outcomes of unobserved latent variables using the common factor model (Harman, 1976). In the common factor model, reflective latent variables are constructed that explain the covariances between the observed indicators (Harman, 1976). As indicated earlier, most researchers have modeled student engagement reflectively. Another option is to specify the construct as a function of observed indicators such that the observations form an underlying latent variable. Principal components analysis is an example of a formative measurement model in which observations are combined to form weighted linear composites intended to represent theoretically meaningful concepts (Joliffe, 2002). Recall the example of SES, in which the latent variable was solely a composite of the observed variables education, income, and prestige (Heise, 1972).

Reflective measurement models are typically used more than formative models in the social sciences and education. However, researchers in a variety of disciplines have made significant effort in making the academic community aware of formative indicators, the potential use of formative measurement models for creating latent constructs, and the consequences of model misspecification when using one type of measurement model versus the other (Bollen & Lennox, 1991; Diamantopoulos, Riefler, & Roth, 2008; Diamantopoulos & Siguaw, 2006; Jarvis, MacKenzie, & Podsakoff, 2003). Although



formative measurement models are gaining traction, Bollen's (1989) statement still remains true, that "most researchers in the social sciences assume that indicators are effect indicators. Causal indicators are neglected despite their appropriateness in many instances" (p. 65). Researchers have proposed two reasons to explain the lack of formative measurement models: 1) many researchers who develop measures may be unaware of the potential appropriateness of formative indicators for operationalizing certain constructs and 2) researchers may avoid specifying formative measurement models because they do not know how to incorporate them into structural equation models (Diamantopoulos et al., 2008).

Studies of student engagement (particularly the NSSE and CCSSE) call into question the appropriateness of using reflective measurement models. McCormick and McClenney (2012) for example have stated that scores from measures of student engagement should not be modeled with confirmatory factor analysis as has been done in the past. It is important to determine the best way to model student engagement prior to drawing conclusions that may have a significant impact on institutional decisions. The following sections compare reflective and formative models, followed by criteria one may employ to choose between the two approaches.

**Reflective measurement modeling.** Many measurement models in the social sciences are reflective (Diamantopoulos et al., 2008; Wang, French, & Clay, 2015). As mentioned earlier, reflective measures are treated as outcomes of a latent variable.

Edwards (2011) depicts reflective measurement using the following equation:

$$x_i = \lambda_i \xi + \delta_i \quad (1)$$

where  $x_i$  is a reflective indicator seen as an item score,  $\xi$  is its associated latent variable,  $\lambda_i$  is the effect of  $\xi$  on  $x_i$ , and  $\delta_i$  is the uniqueness of the measure. A reflective model is also shown in Figure 1a, in which  $\xi$  signifies a latent variable and  $x_1$ ,  $x_2$ , and  $x_3$  are reflective indicators of the construct. The  $\delta_1$ ,  $\delta_2$ ,  $\delta_3$  represent uniqueness associated with the reflective measures and combine item specific variance with random error (Bollen, 1989). The loadings  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  capture the magnitude of the effects of  $\xi$  on  $x_1$ ,  $x_2$ , and  $x_3$ .

The reflective model may be described using an example. Consider a case wherein  $\xi$  could be an employee's perception of their autonomy on his or her job, and  $x_1$ ,  $x_2$ , and  $x_3$  could be scores on the items "I determine the way my work is done," "I have complete control over my schedule", and "I make my own decisions at work". The arrows leading from  $\xi$  to  $x_1$ ,  $x_2$ , and  $x_3$  indicate that respondents' perceived autonomy influences scores on the three items. This premise represents a critical realist perspective (Loevinger, 1957; Messick, 1981), in which constructs are considered real entities that influence scores on their associated measures (Borsboom et al., 2003; Edwards & Bagozzi, 2000). Because all indicators are thought to be the effects of the same latent variable, they are expected to be highly correlated systematically through the latent variable (relating to internal consistency reliability; Bollen, 1984). The indicators should also be interchangeable. That is, the deletion of an indicator should not change the meaning of the latent variable if there are sufficient number of indicators to represent the latent variable.

**Formative measurement modeling.** In contrast to reflective measurement modeling, in formative measurement modeling, constructs are measured with causal

indicators<sup>2</sup>. In other words, observed variables affect levels of the latent construct.

Variables measured with casual indicators can be considered as the combination of multiple observed variables where a change in the indicator affects the underlying latent construct (Bagozzi, 2007; Jarvis et al., 2003). This type of measurement is referred to as formative because the latent variable of interest is in essence formed by the indicators.

The equation for a formative measurement model can be expressed as:

$$\eta = \gamma_i x_i + \zeta \quad (2)$$

where  $x_i$  is a formative measure,  $\eta$  is the construct,  $\gamma_i$  is the effect of  $x_i$  on  $\eta$ , and  $\zeta$  is the residual (i.e., that part of  $\eta$  not measured by  $x_i$ ; Edwards, 2011). A formative measurement model is shown in Figure 1b, where  $\eta$  is the construct of interest and  $x_1$ ,  $x_2$ , and  $x_3$  are formative indicators of the construct. The coefficients  $\gamma_1$ ,  $\gamma_2$ , and  $\gamma_3$  represent the magnitude of the effects of  $x_1$ ,  $x_2$ , and  $x_3$  on  $\eta$ , and the residual  $\zeta$  represents aspects of  $\eta$  not explained by  $x_1$ ,  $x_2$ , and  $x_3$ . Sometimes the residual term  $\zeta$  is excluded from formative measurement models, and thus the latent variable  $\eta$  is an exact weighted linear composite of its measures. Figure 1b shows  $x_1$ ,  $x_2$ , and  $x_3$  freely correlate, and the relationships among the formative measures are indicated by their intercorrelations (Edwards, 2011; MacCallum & Browne, 1993). The indicators in this model may be referred to as *formative* indicators (Bollen & Lennox, 1991). Formative indicators are different from *reflective* indicators in the way in which observed variables reflect the underlying latent factors (Diamantopoulos et al., 2008).

---

<sup>2</sup> “Causal indicators” here refer to observed variables that are assumed to affect the underlying latent variable in formative measurement. Some caution against the use of the term “causal” and question whether measures really cause latent variables. However, the use of the term in the current study comes from Blalock (1964), who was perhaps the first in social and behavioral sciences to call attention to such variables.

An example of formative measurement is socioeconomic status (SES) measured by such observed variables as education, income, and occupational prestige (Heise, 1972). Formative measurement aligns with a constructivist position (Fosnat, 1996). Constructivists view constructs as elements of language in theoretical discourse. Thus, constructs are not attributed any real existence independent of their measurement (Borsboom et al., 2003). The constructs in formative measurement may also be viewed as latent variables that function as analytical devices for combining measures, similar to data reduction in principal components analysis (Borsboom et al., 2003; Edwards, 2011). With respect to SES, it is inappropriate to conceive of SES as a cause of the observed variables, such as occupational prestige. Instead, SES is a function of occupational prestige.

*Identification of formative measurement models.* One of the biggest issues in formative measurement is achieving identification of the model (Edwards, 2011, Wang et al., 2015). Formative models with no reflective indicators or constructs deemed causally subsequent to the formative construct, are not identified models (e.g., Figure 1b). Bollen and Davis (1994/2009) provided a set of rules to determine whether a formative measurement model is identified. As with any latent model, to achieve identification of the model, the number of parameters estimated in a model must be less than or equal to the number of elements in the observed covariance matrix. When the number of parameters equals the number of elements in the observed covariance matrix, this results in a just-identified model. Just-identified models cannot be used to test the fit of the model because they exactly reproduce the observed variances and covariances. For example, the model in Figure 1b is not identified. In order to estimate its parameters, the

model would need to be embedded in a larger model (Bollen & Davis, 1994/2009; Kline, 2011). That is, formative models must include at least two observed or latent variables as outcomes for the purposes of identification. This is known as the *2+ emitted paths rule* and it is necessary but not sufficient to guarantee identification (Bollen & Davis, 1994/2009). Latent variables must also be scaled. Specifically, a scale must be assigned to the latent variable in order for the computer to estimate information about it (Kline, 2011). One of the most common ways of scaling a latent variable is to set the factor loading for one of the latent variable's effect indicators to 1.0 (Kline, 2011). If effect indicators are not included in the model, researchers may set the latent variable's path to another latent variable to one (Bollen & Davis, 1994/2009). Finally, Bollen and Davis (1994/2009) recommend the *exogenous X rule*, which states that each latent variable has at least one observed variable that loads solely onto it and the associated errors are uncorrelated; and each latent variable must have at least two observed indicators in total and the errors of these indicators are uncorrelated with those of the unique indicators.

***Criticisms of formative measurement.*** Formative measurement has been criticized by many for both philosophical and practical reasons. Howell, Breivik, & Wilcox (2007) contended that formative measures are inherently prone to interpretational confounding. They argued that using causal indicators can change the empirical meaning of a latent variable to be something other than that assigned to it by a researcher (Howell et al., 2007). That is, interpretational confounding occurs when:

the assignment of empirical meaning to an unobserved variable ... is other than the meaning assigned to it by an individual *a priori* to estimating unknown

parameters. Inferences based on the unobserved variable then become ambiguous and need not be consistent across separate models (Burt, 1976, p.4).

In a response to this critique of formative measurement, Bollen (2007) asserted that interpretational confounding can occur in formative or reflective models and is a result of misspecification, not the type of indicator used. For example, consider SES is measured by income, education, and occupation. If SES is predicting the outcome of academic achievement, then the nature of SES may become dependent upon the outcome (i.e., academic achievement). Said differently, the nature of SES might change if it is used to predict a different outcome, such as student retention.

Edwards (2011) compared formative and reflective models based on dimensionality, internal consistency, identification, measurement error, construct validity, and causality, and concluded that formative measurement is not practical. He argued that formative measurement is a fallacy because formative constructs cannot be real entities that exist in the world. Edwards maintained that researchers should use alternative models based on reflective measurement that can achieve the same objectives as formative measurement.

Both Edwards (2011) and Howell et al. (2007) suggested designing measures that use reflective indicators. However, a reflective version of a construct may not always denote the same meaning as formative construct. For instance, a person's self-report of their SES tells us their perceived SES, but may not accurately reflect their objective SES. Objective SES could be better modeled formatively using indicators such as income, education, and occupation.

**Choosing a reflective or a formative approach.** Jarvis et al. (2003) provided a set of decision rules for determining whether a construct should be treated as formative or reflective. See Table 1 (adapted from Coltman et al., 2008) for an organizing framework to assess formative and reflective models. The rules fall into four categories: 1) the direction of causality between the construct and the indicators, 2) the interchangeability of the indicators, 3) the covariation between the indicators, and 4) the pattern of the antecedents and consequences of the indicators. Each of the rules is explicated below and a discussion follows as each rule might apply to the measurement of student engagement using the GPI subscale. Chin, Peterson, and Brown (2008) warned researchers that although the criteria proposed by Jarvis et al. (2003) are “intuitively reasonable, ... it is difficult to meaningfully categorize measurement scales unequivocally as being formative or reflective based on the measurement items alone” (p. 289). Although some researchers have conceived of student engagement as a reflective construct (e.g., Campbell & Cambra, 2011; Kuh, 2009; Marti, 2009), consideration of each criteria does not lead to a clear-cut decision that student engagement should be modeled reflectively.

***Direction of causality.*** An important consideration when deciding whether a measurement model is formative or reflective is the direction of causality between the construct and the indicators (Coltman, Devinney, Midgley, & Venaik, 2008; Jarvis et al., 2003). Jarvis et al. (2003) proposed that researchers consider the following questions: “Would changes in the indicators/items cause changes in the construct or not?” or “Would changes in the construct cause changes in the indicators?” (p. 203). In a formative model, changes in the indicators are expected to lead to changes in the level of

the formative construct, whereas the opposite is true in a reflective model. Using this criterion, a student engagement construct measured by different behaviors and perceptions could be considered reflective or formative. For the reflective conceptualization, student engagement could be thought of as a latent “willingness to engage” construct. A person would be expected to increase their behaviors and perceptions in various areas (e.g., participation in the classroom, with their peers) as the willingness to engage latent construct increased. For example, as attachment increases, we might expect responses to the item, “[institution] was the perfect school for me” to increase. Conversely, there is a case for considering these as formative indicators. That is, as a person increases participation with academic related activities and increases perception of the college or university, overall engagement increases. The change in behavior or perception drives the overall change in student engagement, not the other way around. In the formative conceptualization, student engagement is a composite of different demonstrations of engagement. For instance, the item, “My professors at [institution] cared about me as a person” seems to drive the perception of feeling supported by the institution, rather than the perception of support driving the statement.

***Interchangeability of indicators.*** In formative measurement models, indicators do not need to be interchangeable (Jarvis et al., 2003). That is, they do not need to represent similar content. Dropping one indicator in a formative model may therefore alter the definition of the construct. However, in a reflective model indicators are theorized to be chosen from a domain of interchangeable possibilities (Diamantopoulos & Winklhofer, 2001). Considering student engagement measured by the GPI subscale, items related to support, experiential learning, and attachment are not necessarily interchangeable. For



example, the items “While attending [institution], I worked on a project that took a semester or more to complete” and “I was extremely active in extracurricular activities and organization while attending [institution]” are not interchangeable. Although both items relate to experiences during a student’s time at the institution, respondents might answer differently to each item.

***Covariance of indicators.*** In a reflective measurement model, indicators are expected to show moderate to high correlations with one another. High intercorrelations are associated with high reliability as measured by Cronbach’s alpha. In contrast, there is no expectation of internal consistency reliability in formative models (Diamantopoulos & Winklhofer, 2001). In a formative model, high inter-item correlations could be problematic because each item is expected to uniquely contribute to the latent variable. We would not necessarily expect to see moderate to high correlations across different domains of student engagement (e.g., high correlations between experiential learning and attachment). We might expect some behavioral components to be correlated with one other, but they may not be correlated with the perceptual aspects of engagement. For example, it is plausible to think that working on a research project and having an internship/job where a student applied what they learned in the classroom might be related. However, we might not expect working on a research project to be related to whether someone feels the institution was the perfect place for them. Using correlations as a criterion to distinguish formative from reflective models, it is possible that aspects of student engagement might fit either a formative or a reflective measurement model.

***Antecedents and consequences of indicators.*** Formative indicators do not need to have the same antecedents and consequences as one another (Jarvis et al., 2003). The

items are not necessarily interchangeable and may capture different aspects of the construct's domain (Jarvis et al., 2003). For example, education, income, and occupational prestige (i.e., measures of SES) are not interchangeable and potentially have different antecedents and consequences. Variables that lead to increased income do not necessarily lead to more education. In the proposed model of student engagement, the behavioral and perceptual aspects of engagement may have different antecedents. For example, participating in a research or academic activity depends on the availability of such activities; having a faculty member as a mentor depends on opportunity and time for such a relationship to occur. That the indicators may have different antecedents, makes the case for a formative model.

### **Modeling Student Engagement in the Current Study**

The current study had two primary areas of focus: 1) examining validity evidence of the GPI-SEAS items and 2) determining which measurement modeling approach (reflective or formative) best captured theoretical and empirical conceptualizations of student engagement. The differing results and methods of modeling student engagement in previous studies and the lack of psychometric evaluation of the GPI student experiences and attachment subscale (GPI-SEAS) motivated the current study. Recall, the GPI-SEAS includes 10 items grouped into three categories: experiential learning, support, and attachment.

**Evaluation of the GPI-SEAS using Benson's Framework and Jarvis et al.'s Criteria.** Benson's (1998) three-stage construct validation framework includes: 1) a substantive stage that is concerned with a clear definition of the theoretical and empirical domains of student engagement; 2) a structural stage focused on the dimensionality of the

GPI-SEAS; and 3) an external stage that emphasized the relationship between student engagement and other constructs, using structural equation modeling.

Little information is known about how Gallup defined the areas of experiential learning, support, and attachment included in the GPI-SEAS. The items in each of the three categories on the GPI-SEAS have wording and categories similar to those on other student engagement instruments, such as the NSSE. However, it is unclear whether the creators of the Gallup instrument used theory, previous research, or content experts to develop the GPI-SEAS items. A discussion of various student engagement definitions can be found earlier in this chapter. Recall, within the context of this study, student engagement is defined as students' involvement in academic and co-curricular experiences provided by the institution consisting of affective, cognitive, and behavioral investment. The GPI-SEAS includes items related to the affective and behavioral (but not cognitive) aspects of student engagement.

Because of the lack of background information on the creation and inclusion of the GPI-SEAS, Jarvis et al.'s (2003) criteria was used to examine the item content and discuss the structural stage of content validity evidence. Specifically, using the direction of causality criterion (Jarvis et al., 2003), it seems plausible to measure some items formatively and others reflectively. That is, student engagement could be thought of as a latent "willingness to engage" or it could be thought of as a composite of different demonstrations of engagement. Considering the interchangeability of indicators criterion (Jarvis et al., 2003), it does not seem as though the indicators on the GPI-SEAS are interchangeable. Namely, items such as "While attending [institution], I worked on a project that took a semester or longer to complete." and "My professors at [institution]

cared about me as a person.” do not necessarily represent the same content. Having a professor who cares may in fact cause attachment as opposed to attachment causing responses to the item. The inappropriateness of interchangeability indicates that the construct should be measured formatively or that the items represent separate factors of student engagement.

The covariance of indicators criterion (Diamantopoulos & Siguaaw, 2006) asks whether it is necessary for the indicators to covary. Thus, depending on intercorrelations of the items, we might also model the items based on the behavioral and affective components of student engagement. For example, we might expect the experiential learning items to covary with one another, but not to covary with the attachment items. Therefore, based upon the covariance of indicators criterion, it makes sense to measure each of the three areas (experiential learning, support, and attachment) as separate reflective models. However, we might also expect some of the support and attachment items to covary because they both reflect affective components of engagement. Using the antecedents and consequences of indicators criterion (Diamantopoulos & Siguaaw, 2006; Jarvis et al., 2003), different aspects of student engagement may have different antecedents (e.g., participating in extracurricular activities versus having a faculty mentor), which makes the case for using a formative model.

In the current study, I focused on providing appropriate evidence of validity for the second two stages of Benson’s (1998) framework using both reflective and formative approaches. The lack of information for the substantive stage is noted in the Limitations for the current study.

To provide support for the structural stage (Benson, 1998), exploratory factor analysis and confirmatory factor analysis were used to model the GPI-SEAS items reflectively. That is, the GPI-SEAS is split into three different categories, which suggests a three-factor structure is plausible for the GPI-SEAS. The three-factor structure hypothesis was compared to two competing hypotheses including a one-factor model, which represents a unidimensional student engagement factor and a two-factor model, which represents the affective and behavioral aspects of student engagement.

Establishing the structural stage provides necessary, but not sufficient evidence of construct validation (Nunnally, 1978). Benson (1998) considered the final stage of external validation to be the “most crucial” because it involves what is actually being measured by the GPI-SEAS (p. 14). For the GPI-SEAS, Benson’s external stage is concerned with the relationship between the hypothesized structure of the student engagement construct of the GPI-SEAS and other observed variables. Structural equation modeling has been suggested as an ideal method to study the external stage of construct validation (Benson, 1998).

To provide evidence for Benson’s external stage, the relationship between GPI-SEAS items and life satisfaction and work satisfaction was examined. To gather validity evidence for the reflective model, I estimated a full structural theoretical model linking the GPI-SEAS to the observed variables of life satisfaction and work satisfaction. To gather validity evidence for the formative model, I examined the magnitude of the direct structural relationships between the indicators and the latent variable as well as the variance of the latent variable (Bollen, 2011).

## The Current Study

It is apparent that student engagement is a complex concept. Modeling student engagement using the GPI-SEAS requires knowledge of student engagement theory and determination of whether the construct should be measured formatively or reflectively. An evaluation of Jarvis et al.'s (2003) decision criteria aid in making a case for modeling some aspects of the GPI-SEAS formatively and others reflectively. For this reason, alternative models using both formative and reflective approaches were employed and each model was evaluated to determine which structure and conceptualization of student engagement should be supported. Ultimately, six research questions were posed and were answered across three study stages. In the first stage, inter-item correlations among the 10 GPI-SEAS items were examined and a reflective measurement model was estimated using exploratory factor analysis. In the second stage, confirmatory factor analysis was conducted to see if there was support for a reflective measurement model. Finally, in stage three, formative models of the GPI-SEAS were estimated using life satisfaction and work satisfaction as outcome indicators. The choice between a reflective and formative model cannot solely be answered via empirical evidence. Thus, the study is guided by both empirical and logical/philosophical questions to determine which type of model to champion. The research questions were:

**Stage I: Item analysis and estimating reflective models using exploratory factor analysis.** In the first stage of the study, I examined descriptive statistics and inter-item correlations using an independent subsample of the data. In this stage, I also conducted an exploratory factor analysis. The research questions for this stage were:

1. Does the pattern of inter-item correlations for the GPI-SEAS suggest a one-, two-, or three-factor model?
2. What is the number and nature (i.e., pattern of loadings) of the factor(s) that may account for the set of inter-item correlations?

**Stage II: Estimating reflective models using confirmatory factor analysis.** In the second phase of the study, I further investigated whether a reflective model represented the data. Unidimensional (i.e., one general student engagement factor), two-factor (representing the behavioral and affective aspects of student engagement), and three-factor (representing support, experiential learning, and attachment) nested CFA models were estimated using the 10 GPI-SEAS items. In this stage, I gathered external validity evidence for the championed model by examining its relationship with theoretically-related external variables. Specifically, the life and work satisfaction outcome variables were added to the CFA model, similar to a full hybrid structural equation model. The research questions for this stage were:

3. What is the dimensionality of the GPI-SEAS using reflective models guided by relevant student engagement literature? Does a hypothesized one-, two-, or three-factor model best represent the data?
4. What is the magnitude and direction of the relationship between student engagement, life satisfaction, and work satisfaction when measured reflectively?

**Stage III: Estimating formative models.** In the third phase of the study, I examined whether formative models represented the data. The research questions included specific criteria discussed by Bollen (2011) for examining formative models. The research questions for this stage were:

5. How well does a formative model represent the GPI-SEAS items?
6. What is the magnitude and direction of the relationship between student engagement, life satisfaction, and work satisfaction when measured formatively?



## CHAPTER 3

### Methods

Data for the current study were collected from alumni who attended a mid-sized southeastern university and graduated with a bachelor's degree between 1996 and 2005. Descriptions of the measures are presented first, followed by a description of the sample and data collection procedures. A description of data analyses completes this section.

### Measures

**GPI Student Experiences and Attachment Scale (GPI-SEAS).** The GPI-SEAS (Gallup-Purdue, 2014) is a self-report measure developed to assess alumni perceptions of their undergraduate experience and their attachment to the institution. The ten student experiences and attachment items are intended to reflect experiential learning (e.g., “While attending [institution], I worked on a project that took a semester or more to complete”), support (e.g., “My professors at [institution] cared about me as a person”), and attachment to the institution (e.g., “[institution] was the perfect school for people like me”; Gallup-Purdue, 2014). There are three items related to experiential learning, three items related to support, and four items related to attachment (see Table 2). In addition, the three experiential learning items were hypothesized to represent the behavioral aspect of student engagement, and the support and attachment items were hypothesized to represent the affective aspect of student engagement. Response options ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). Currently, there is no published evidence of reliability or validity evidence for the subscales.

**Cantril Self-Anchoring Striving Scale (Cantril Scale).** The Cantril Scale (Cantril, 1965) was developed to assess a type of well-being related to judgments of life

or life evaluation (Diener, Kahneman, Tov, & Arora, 2009). The instrument asks respondents to rate where they currently stand on a ladder that represents the best or worst possible life for them. Response options range from 0 (*worst possible*) to 10 (*best possible*). The instrument includes an item related to the present (i.e. *On which step of the ladder would you say you personally feel you stand at this time?*) and an item related to the future (i.e. *On which step do you think you will stand about five years from now?*). Only the item asking about the “present” was used for this study as an indicator of life satisfaction. That is, where respondents thought they would be satisfied with their lives in the future was not of interest in the current study.

**GPI Workplace Satisfaction Item (GPI-WSI).** The GPI-WSI is a self-report item of work satisfaction. The item used for the current study stated, “I am deeply interested in the work that I do.” Response options ranged from 1 (*strongly disagree*) to 5 (*strongly agree*).

### **Participants**

All data analyzed for this study were collected from a total of 1,340 alumni who graduated from the institution between 1996 and 2005. All data were collected during spring 2016. Of the participants who responded to the survey, 91% were white and 55% were female. Only cases with complete data on all GPI-SEAS items and the Cantril Scale and GPI-WSI items were retained in the analyses.

### **Data Collection Procedures**

Participants were identified via an alumni database and initially 10,500 undergraduate alumni were contacted via email. Alumni were recruited through a series of four emails: a preliminary informational email sent from the institution, the first survey

elicitation email, and two follow-up emails requesting participation. Emails were crafted jointly by research team members at the institution and Gallup. Of the 10,500 alumni who were solicited for participation, 1,340 responded to the survey (12.8% response rate). The GPI-SEAS, the Cantril Scale, and the GPI-WSI (and several additional subscales) were administered as part of the survey.

### **Data Analyses**

The study was conducted in three stages. First, the sample was randomly split into two independent subsamples prior to the analyses. After removing the 291 cases that contained missing data, 1,049 alumni comprised the sample that was used for this study (dropping the response rate to 10%). Of these respondents, 91.1% were White and 53.6% were female, which is representative of the larger population at the institution. Therefore, the sample of respondents reflect the race of the overall student population, but males are over-represented in the sample used for the current study.

The first independent subsample included 349 randomly selected respondents and the second subsample included the remaining 700 respondents. The purpose of Stage I was to complete principal axis factor analysis using data from independent subsample one. In Stage II, I continued to examine the internal structure of the GPI-SEAS using reflective measurement models by testing a series of nested confirmatory factor analysis models. The second independent subsample was used for the analyses in Stage II. In Stage III, also using independent subsample two, formative models of the GPI-SEAS were estimated to evaluate whether they fit the data (see Table 2)

SPSS version 23.0 (IBM Corporation, 2014) was used to randomly divide the data into two independent subsamples, calculate descriptive statistics, and perform the item

analysis. Exploratory factor analysis was conducted using the EFA standalone software package, FACTOR (Lorenzo-Seva & Ferrando, 2006) which was designed for the use of ordinal data. Mplus version 7.3 (Muthén & Muthén, 2014) was used to screen the data and complete all analyses in Stages II and III involving structural equation modeling.

**Stage I: Item analysis and reflective exploratory factor analysis using the first independent sample.** In Stage I of this study, I addressed the first two research questions. Specifically, these questions were:

1. Does the pattern of inter-item correlations for the GPI-SEAS suggest a one-, two-, or three-factor model?
2. What is the number and nature (i.e., pattern of loadings) of the factor(s) that may account for the set of inter-item correlations?

Prior to the analyses for Stage I, descriptive statistics and inter-item polychoric correlations were analyzed for subsample one (Table 3). Inter-item correlations were examined to determine if they fit with theoretical hypotheses, for a one-, two-, or three-factor model, as outlined previously in chapter two. Specifically, the hypotheses included a one-factor general student engagement model, a two-factor model of student engagement representing the behavioral (Items 8-10) and affective (Items 1-7) aspects of the construct, and a three-factor model representing the support (Items 3, 4, and 6), experiential learning (Items 8-10), and attachment (Items 1, 2, 5, and 7) areas posited by Gallup (2014). See Table 2 for a detailed alignment of the items for the two-and three-factor models.

Given data on the GPI-SEAS were on a five-point Likert type scale and we cannot be sure that there are equal intervals between each scale point, data were treated as

ordered categories. Conventional EFA is based on the Pearson correlation matrix. Pearson correlations assume data have been measured on an equal interval scale and a linear relationship exists between the variables. Because data for the current study were treated as ordinal, Pearson correlations were not appropriate. Compared to polychoric correlations, Pearson correlations have been found to underestimate the strength of relationships between ordinal items (Olsson, 1979). The polychoric correlation matrix was used for the EFA analyses in this study.

Parallel analysis based on the minimum rank factor analysis (PA-MRFA; Timmerman & Lorenzo-Seva, 2011) and percent of common variance were examined as extraction criteria for the exploratory factor analysis. Horn's (1965) parallel analysis and the scree plot (Cattrell, 1966) are commonly recommended methods of factor extraction. PA-MRFA, which is based on random permutation of the sample data and comparing the percentage of common variance extracted by MRFA, was found to outperform Horn's parallel analysis (Baglin, 2014). The scree plot has been shown to overestimate the number of dimensions in the data (Ruscio & Roche, 2012). Therefore, the scree plot was not examined in the current study.

Distinct solutions were examined for both an approximation of simple structure and for theoretical meaningfulness. The percentage of common variance explained from the randomly permuted data to the observed explained common variance was analyzed to determine how many factors to extract. The resulting number of factors was then extracted and rotated to a final solution using a direct oblimin rotation

**Stage II: Estimating reflective CFA models using the second independent sample.** In the second stage of the study, I examined the reflective models of the GPI-SEAS using confirmatory factor analysis. Stage II addressed two research questions:

3. What is the dimensionality of the GPI-SEAS using reflective models guided by relevant student engagement literature? Does a hypothesized one-, two-, or three-factor model best represent the data?
4. What is the magnitude and direction of the relationship between student engagement, life satisfaction, and work satisfaction when measured reflectively?

The single factor model representing a general student engagement factor and the two-factor model representing the behavioral and affective components of student engagement supported by various researchers (e.g., Chapman, 2003; Fredericks et al., 2004; McCormick et al., 2013; Skinner & Belmont, 1993) were tested (See Figures 2 and 3). The three-factor model representing the three areas (i.e., experiential learning, support, and attachment) purported by Gallup-Purdue (2014) was also tested (See Figure 4). Given the three hypothesized models were nested, the models were compared (explained in greater detail later in this chapter) to determine whether one model fit statistically significantly better than the other. To obtain additional validity evidence, a full structural model was then estimated including the championed CFA model and life satisfaction and work satisfaction as outcome variables.

***Data screening and assumptions.*** Again, because the data were treated as ordinal, the polychoric matrix for subsample two (Table 4) was used for structural equation modeling analyses. Prior to conducting any analyses, the following data screening procedures were conducted.

First, data were screened for univariate and multivariate outliers. Data on each individual variable were examined for extreme scores to identify univariate outliers. To identify multivariate outliers, a regression procedure was used to obtain the Mahalanobis distance (the distance of a case from the centroid in a multivariate space). A break in the list of the top ten Mahalanobis distance values was used to detect multivariate outliers. Values identified as both univariate and multivariate outliers were examined closely to assess if their responses were anomalies and to determine if they should be excluded from the analyses.

Second, data were screened for both univariate and multivariate normality. Absolute values above  $|2|$  for skew and above  $|7|$  for kurtosis were used to detect univariate non-normality (Chou & Bentler, 1995; Finney & DiStefano, 2013). To screen for multivariate non-normality, Mardia's normalized multivariate kurtosis was used as a criterion. There is no strict cutoff value for Mardia's coefficient. However, Bentler (1998) indicated in a SEMNET post that values of 10 or 20 are a "pretty good indication that the data truly are not normal."

***Structural equation modeling.*** Structural equation modeling (SEM) is a family of statistical techniques that can be used with reflective or formative models. SEM allows for exploratory and confirmatory modeling of the complex relationships among latent variables, their observed indicators, and additional observed variables (Bollen, 1989). The use of SEM requires the researcher to develop a theoretically-based view of the phenomenon of interest. Models are specified a priori (generated prior to looking at the data). SEM techniques provide the researcher with statistics (i.e., chi-square tests and goodness-of-fit indexes) in order to evaluate global and local fit of the model (Kline,

2011). The purpose of SEM is to develop a theoretically defensible empirical model. The researcher should question whether the model “works to achieve its goals... compared with other models that are reasonable competitors” (Rodgers, 2010, p. 4). The researcher should also be encouraged when using SEM to consider a range of competing models, based on comprehensive theoretical and practical understanding of the subject matter.

A structural equation model includes both a measurement model (the factor model) and a structural portion (similar to the path model, but with relationships specified among the latent rather than the observed variables). In the commonly used two-step modeling approach, the measurement part of the model is examined using a confirmatory factor analysis (CFA) model to analyze the relationship between the indicators and the latent variables (Anderson & Gerbing, 1988; Bollen, 1989). The measurement model is estimated to see if good fit is achieved and if so, then the structural model is evaluated.

***Model identification.*** A crucial step in latent variable modeling is ensuring that the specified model is identified, and preferably, overidentified. A model is identified if it is possible to arrive at a unique estimate for each free parameter (path coefficients, variances, and covariances) in the model (Kline, 2011). Structural equation models comprise a set of linear equations estimated using observed variances and covariances. The first requirement for identification is there must be more observations than parameters to be estimated (Kline, 2011). If there are more parameters than observations, it will be impossible to find unique estimates for the parameters. An identified model can be *just-identified* or *overidentified*. A just-identified model has the same number of parameters as observations. If the model meets the other requirement of having a scale assigned to each latent variable, it can be estimated but model fit cannot be evaluated



(Kline, 2011). Generally, researchers evaluate overidentified models, which have fewer free parameters than observations. The second requirement is that each latent variable must be assigned a scale (metric) so estimates involving the latent variables can be calculated (Kline, 2011).

*Estimation method.* The estimation method should provide the most efficient, unbiased, and consistent parameter estimates given the ordinal nature of the data in the current study. It was inappropriate to analyze the data using typical theory estimators or Pearson correlation coefficients, because GPI-SEAS scores were not continuous or normally distributed. That is, an estimator such as Maximum Likelihood (ML) is not appropriate for ordinal data because it assumes continuous data that follows a multivariate normal distribution. Analyzing non-normal ordinal data using ML estimation produces biased standard errors,  $\chi^2$  values, and parameter estimates (Finney & DiStefano, 2013).

Robust Diagonally Weighted Least Squares (RDWLS) was used to estimate the hypothesized models because it accounts for the ordinal nature of the data by analyzing polychoric correlations rather than Pearson correlations or covariance matrices. Similar to WLS estimation, RDWLS uses the asymptotic covariance matrix of the polychoric correlations as the weight matrix. However, RDWLS estimation does so without having to invert the full asymptotic covariance matrix (Finney & DiStefano, 2013). Specifically, RDWLS estimation only requires inverting the diagonal of the asymptotic covariance matrix (Finney & DiStefano, 2013). When models are complex or sample sizes are small, RDWLS outperforms WLS because RDWLS does not have to invert the full asymptotic covariance matrix. Unbiased and consistent parameter estimates are produced using only

a portion of the asymptotic covariance matrix. However,  $\chi^2$  statistics and standard errors will be biased. Thus, the DWLS  $\chi^2$  test statistic and standard errors must be adjusted using information from the full asymptotic covariance matrix. The RDWLS estimator uses scaling techniques similar to those used in the Satorra-Bentler scaling procedure (Satorra-Bentler, 1994), to adjust the DWLS  $\chi^2$  statistic and biased standard errors of the parameter estimates (Finney & DiStefano, 2013). Similarly, fit indices are modified using the adjusted (robust)  $\chi^2$  statistic.

*Evaluation of model-data fit.* To evaluate model-data fit, local and global fit were analyzed. Local misfit was assessed by examining polychoric correlation residuals for pairs of items. In addition, global fit was assessed by using a few commonly-used fit statistics and indexes. Fit indices can be categorized as absolute or incremental. Absolute fit indexes evaluate how well the model reproduces the sample data in an overall sense. Incremental fit indexes evaluate the fit of a hypothesized model relative to a baseline model (Hu & Bentler, 1999). The following fit indices were used for the current study.

*Chi-square goodness-of-fit test ( $\chi^2$ ).* The DWLS adjusted  $\chi^2$  statistic is an absolute index of exact fit. That is, the DWLS adjusted  $\chi^2$  statistic tests how well the model *exactly* fits the data (Weston & Gore, 2006). A significant  $\chi^2$  value means the researcher should reject the null hypothesis that the model perfectly fits the data (Kline, 2011). Non-significant chi-square values indicate the model may reasonably represent the data. The model chi-square statistic has some limitations. The null hypothesis of perfect fit is unrealistic and unlikely to ever be supported (Kline, 2011). The chi-square statistic is also sensitive to sample size, and as sample size increases this index may cause a researcher to incorrectly reject the model (Bentler & Bonnett, 1980; Kline, 2011). Issues with the

model  $\chi^2$  statistic led SEM methodologists to develop a variety of absolute and incremental fit indexes.

*Standardized root mean square residual (SRMR).* The SRMR is an absolute fit index that is based on the residuals between the elements of the observed and model implied covariance matrices (Kline, 2011). Hu and Bentler (1998) recommend this index always be reported to assess model fit. SRMR ranges from zero to one and values less than .08 can be used to indicate adequate fit (Hu & Bentler, 1999).

*Root mean square error of approximation (RMSEA).* The RMSEA was evaluated to assess absolute fit. RMSEA is sensitive to parsimony. That is, simpler models (with fewer degrees of freedom) will have lower RMSEA values. Values of RMSEA range from zero to one, with values closer to zero indicating better model-data fit. Based on a simulation study of CFA models, Hu and Bentler (1999) proposed that an RMSEA cutoff around .06 is appropriate. Other common guidelines are to consider RMSEA less than or equal to .05 as good fit, RMSEA between .05 and .08 as adequate, and values of .10 or greater as indicators of poor fit (Browne & Cudeck, 1993).

*Comparative fit index (CFI).* In addition to absolute fit indexes, Hu and Bentler (1998, 1999) also recommend reporting incremental fit indexes in SEM studies. The CFI is an incremental fit index that represents the proportion of improvement in fit of the hypothesized model over the baseline model. It is scaled to have a value between zero and one, with higher values representing better model fit (Kline, 2011). Hu and Bentler's (1999) simulation study suggested values of .95 or greater to indicate adequate fit.

*Evaluation of model parameters.* Estimated parameters for adequately fitting models were interpreted. Both standardized and unstandardized pattern coefficients,

standard errors,  $z$ -tests, and  $p$ -values for every estimated parameter were reported and interpreted. In addition, the amount of variance explained in each factor of interest was reported and interpreted. For each latent factor in the CFA analyses, McDonald's (1999)  $\omega$  was calculated using unstandardized parameter estimates and error variances. Values at or above .70 are considered acceptable (McDonald, 1999). McDonald's  $\omega$  is considered a more accurate estimate of internal consistency than Cronbach's alpha because it allows only the variance due to the factor of interest to be treated as systematic variance (McDonald, 1999). Means, standard deviations, and inter-factor correlations for the resultant factors will also be reported.

*Nested model comparisons.* Stage II involved an examination of alternative nested models. Although a more complex model always fits the data better than a less complex model, it is important to evaluate whether the difference is significant enough to justify championing a more complex, but less parsimonious, model (Steenkamp & Baumgartner, 1998). In the three-factor model, one factor represented the "attachment" items (Items 1, 2, 5, and 7), the second factor represented the "support" items (Items 3, 4, and 6), and the third factor represented the "experiential learning" items (Items 8, 9, and 10). In the two-factor model, the "support" and "attachment" items (Items 1-7) were combined to represent an "affective" factor. The "experiential learning" items from the three-factor model represented a "behavioral" factor in the two-factor model. The one-factor model included all 10 GPI-SEAS items to represent a general student engagement factor. Table 2 includes details of the alignment of the GPI-SEAS items for the two- and three-factor models.

To compare the relative fit of one-, two-, and three-factor models in the structural equation modeling framework, chi-square difference tests were used to determine if one model fits statistically significantly better than others. The chi-square difference test is a test of statistical significance. Therefore, if chi-square values are statistically significantly different, this will indicate one model fits better than the other. See Figures 2, 3, and 4 for depictions of the competing hypothetical models of interest.

***Full structural model.*** Once adequate model fit was achieved for the reflective models and a model was championed, then a full structural model was estimated including life satisfaction and work satisfaction as outcomes. That is, in the full structural model, the factors from the championed model influence life satisfaction and work satisfaction (See Figure 5). Specifically, I hypothesized that attachment and support would influence the life satisfaction outcome and support and experiential learning would influence the work satisfaction outcome (See Figure 6). These hypotheses were based on previous research that showed positive relationships between a “support campus environment” benchmark (on the NSSE) and a life satisfaction outcome measured one to two years after graduation (Kuh, 2009; Schmaling & Guy, 2014). Prior research on the Gallup-Purdue Index showed that respondents who participated in experiential learning activities and who felt supported by the institution, were more engaged and satisfied in their jobs than respondents who may not have had those experiences (Ray & Kafka, 2014).

Latent variables were scaled using the metric of the indicators. Fit indices along with standardized and unstandardized coefficients were analyzed for the full model. Polychoric correlation residuals as well as statistical significance of the paths were also

reported and interpreted. Fit of the models is discussed in Chapter 4 in comparison to the championed CFA model (based on chi-square difference tests) and in terms of parsimony and relationships according to theory.

**Stage III: Estimating the formative models using the second independent sample.** In the third stage of this study, I examined formative models of the GPI-SEAS using MIMIC-style models. Stage III addressed two research questions:

5. How well does a formative model represent the GPI-SEAS items?
6. What is the magnitude and direction of the relationship between student engagement, life satisfaction, and work satisfaction when measured formatively?

Three formative measurement models of student engagement were estimated based on latent variables formed 1) from all indicators to represent a “student engagement” latent variables (see Figure 7); 2) from the items grouped by behavioral or affective aspects of student engagement (see Figure 8); and 3) from the three areas explained by Gallup-Purdue (2014; see Figure 9): experiential learning, support, and attachment (see Figures 7, 8, and 9). The formative indicators that were grouped together in each model were allowed to covary as recommended by MacCallum and Browne (1993) and Jarvis et al. (2003).

**Model identification.** As Bollen and Davis’ (1994/2009) 2+ emitted paths rule states, a formative construct is underidentified unless it is embedded in a model in which it includes two outgoing paths. The formative model included two reflective indicators, life satisfaction and workplace satisfaction. Bollen and Davis (1994/2009) maintain that the 2+ emitted paths rule is a necessary, but not sufficient condition for identification. In

addition, I will define a scale for the latent variable in the formative model by setting a path (to one of the indicators) to one.

*Evaluation of model-data fit.* To evaluate the appropriateness of the formative models to the GPI-SEAS, global fit indices were first examined. Like the reflective CFA models, the adjusted DWLS  $\chi^2$  statistic, RMSEA and CFI fit indices were evaluated. SRMR was not provided for MIMIC models in Mplus using DWLS estimation. However, Yu & Muthén (2002) suggest that CFI and RMSEA perform well with ordinal data. In addition to the fit indices, the validity of each formative indicator was examined. Bollen (2011) provides guidance on how to evaluate the formative indicators. Specifically, he recommends assessing the path coefficients, the variance explained in the latent construct, and multicollinearity. Standardized and unstandardized coefficients were examined to determine whether they were statistically significant. In addition, I analyzed the magnitude of variance that was explained for the latent variable. Finally, multicollinearity among the indicators was examined. Ideally, the indicators should have low collinearity, otherwise it is difficult to estimate their individual effects (Bollen, 2011).

## CHAPTER 4

### Results

Recall that the current research study consisted of three stages with research questions subsumed under each stage. Stage I included an item analysis and exploratory factor analysis of the GPI student experiences and attachment scale (GPI-SEAS). Stage II included an examination of the internal structure of the GPI-SEAS items using reflective CFA measurement models. Specifically, a series of three nested models were tested using confirmatory factor analysis. Finally, in Stage III, formative models of the GPI-SEAS items were estimated. The formative models included two reflective indicators (i.e., life satisfaction and work satisfaction) for the purposes of identification. Below, the results for each research question are presented, followed by a general discussion of all findings. A summary of results for each stage can be found in Table 5.

#### **Stage I: Item Analysis and Estimating Reflective Models using Exploratory Factor Analysis**

**Research question 1.** Does the pattern of inter-item correlations for the GPI-SEAS suggest a one-, two-, or a three-factor model? Descriptive statistics and inter-item polychoric correlations for Subsample 1 ( $n = 349$ ) are presented in Table 3. The polychoric correlations ranged from .111 to .722. All polychoric correlations were statistically significant at  $p < .05$ .

Examination of the inter-item polychoric correlations showed some evidence for a two-factor solution, as the “support” and “attachment” items had higher inter-item correlations with one another than with the experiential learning items. However, the polychoric correlations did not show clear evidence for a three-factor solution grouped by



“support”, “attachment”, and “experiential learning” items. That is, a few of the attachment and support items were moderately correlated. For example, Item 1 (attachment) and Item 3 (support) had an observed polychoric correlation of .569. Additionally, Item 4 (support) and Item 5 (attachment) had an observed polychoric correlation of .600. Again, this provided possible support for combining the support and attachment items into an “affective” factor (leaving the “experiential learning” items to represent a “behavioral” factor in the hypothesized two-factor model).

The three “experiential learning” items had relatively low inter-item polychoric correlations with one another, ranging from .182 to .311. The low correlations among items that were expected to group together signaled potential issues for the factor analysis. Based on the polychoric correlations, it was plausible that the “experiential learning” items would not group together as a distinct factor, especially because these items had slightly higher correlations with other items. Based on the polychoric correlations of the “experiential learning” items, a one-factor solution was also plausible.

**Research question 2.** What is the number and nature (i.e., pattern of loadings) of the factor(s) that may account for the set of inter-item correlations? To address this research question, EFA was conducted for Subsample 1 using the free, standalone EFA package FACTOR, which was designed for the use of ordinal data (Lorenzo-Seva & Ferrando, 2006). Data were analyzed via parallel analysis based on the minimum rank factor analysis (PA-MRFA; Timmerman & Lorenzo-Seva, 2011) and percent of common variance were examined as extraction criteria for the exploratory factor analysis. Distinct solutions were also examined for both an approximation of simple structure and theoretical meaningfulness. An examination of extraction criteria led to a one-factor

solution. However, based on the equivocal inter-item polychoric correlations, EFA was conducted specifying one-, two-, and three-factor solutions. Neither the inter-item correlations nor the extraction criteria clearly supported a three-factor solution, which was the hypothesized model based on Gallup's (2014) distinction of the items as support, attachment, and experiential learning.

When conducting PA-MRFA, five-hundred randomly permuted matrices that consisted of the same number of people and variables were created. PA-MRFA is based on the random explained common variance, rather than eigenvalues (Baglin, 2014). The FACTOR program uses two criteria to make the decision regarding how many factors to retain. One criterion is based on the mean of random variance extracted and the other is based on the 95<sup>th</sup> percentile of random percentage of variance (Baglin, 2014). The mean or 95<sup>th</sup> percentile of the factor's percentage of common variance is compared to the observed explained common variance from the sample. If a factor's observed percentage of explained variance is greater than the random percentage, then the factor is retained. This only occurred for the first factor in the parallel analysis (Table 6). That is, the observed data percentage of variance for the first factor, 51.0, exceeded the 95<sup>th</sup> percentile of the random common variance extracted, 23.9. For the second factor, the common variance from the 95<sup>th</sup> percentile of random variance, 17.6, was greater than the observed data percentage, 12.8. Therefore, one factor was retained. Additionally, the one-factor solution explained approximately 68 percent of the common variance.

A one-factor model was estimated representing a general student engagement factor. Table 7 displays the factor loadings for the one-factor solution. Pattern and structure coefficients  $> .40$  were considered salient. Based on the extraction criteria, it

appeared that a one-factor model offered the most parsimonious solution. That is, nine of the ten items had structure coefficients  $> .40$  (ranging from .443 to .749). Item 9 (“While attending [institution], I worked on a project that took a semester or more to complete”) had a structure coefficient of .303. If the GPI-SEAS truly has a unidimensional factor structure, Gallup’s three distinct categories (i.e., support, attachment, and experiential learning) should be reconsidered.

Recall, student engagement also has affective and behavioral aspects. Therefore, a two-factor solution seemed plausible and was also tested. Theoretically, the “attachment” and “support” items would form a factor and the experiential learning items would form a separate factor. This model also seemed plausible based on the polychoric correlations, in which several of the “attachment” and “support” items showed moderate to high inter-item correlations with one another. As seen in Table 7, after direct oblimin rotation was applied, most the items did not have salient loadings with what theoretically should have been behavioral or affective aspects of engagement. Items 1-3 had structure and pattern coefficients  $> .40$  on one factor, while Items 4-9 (a mix of all three item types) showed structure and pattern coefficients  $> .40$  on the second factor. Item 10 (“I was extremely active in extracurricular activities and organizations while attending [institution]”) had similar structure coefficients for both factors (.36 for factor one and .41 for factor two). However, both pattern coefficients for Item 10 were  $< .40$ . The first and second factor were moderately correlated at .511. It does not appear that the two-factor solution represents the affective and behavioral aspects because the two factors were not theoretically interpretable.

Each item on the GPI-SEAS was deemed by Gallup-Purdue (2014) to align with support, attachment, or experiential learning. Therefore, theoretically, each item would show structure and pattern coefficient  $> .40$  on the factor that corresponds with its grouping based on Gallup's designation. Factor loadings for the three-factor solution are presented in Table 7. The three-factor model did not result in interpretable factors. For example, items specified as "attachment" did not "hang together" more with each other than with items specified to measure "support" (see Table 7). Specifically, Items 1, 2, and 3 (a mix of support and attachment items) had coefficients  $> .40$  on factor two. Items 5-7 (again, a mix of support and attachment items) had coefficients  $> .40$  on factor one. Item 8 was the only item with a coefficient  $> .40$  on factor three. Items 9 and 10 failed to have coefficients  $> .40$  on any factor. Factor one was moderately correlated with factor two ( $r = .514$ ) and factor two ( $r = .448$ ), and factor two had a small correlation with factor three ( $r = .265$ ). The three-factor solution does not appear to represent the "support", "attachment", and "experiential learning" categories detailed by Gallup-Purdue (2014).

The EFA results seemed to support a one-factor structure for the GPI-SEAS. Based on the EFA results, the next logical step was to test the internal factor structure of the GPI-SEAS using confirmatory factor analysis.

## **Stage II: Estimating Reflective Models using Confirmatory Factor Analysis**

**Research question 3.** What is the dimensionality of the GPI-SEAS using reflective models guided by relevant student engagement literature? Does a hypothesized one-, two-, or three-factor model best represent the data? Recall that answering research question 3 required estimating three nested CFA models, including a one-, two-, and three-factor model. The three models that were tested included: 1) a one-factor general

student engagement model; 2) a two-factor model consisting of an affective factor and behavioral factor; and 3) a three-factor model consisting of a support factor, an attachment factor, and an experiential learning factor.

**Data screening and guidelines for evaluating model fit.** Prior to the analyses, the data were screened for outliers and non-normality. Based on Mahalanobis distances, no multivariate outliers were removed for the analyses. Means, standard deviations, skewness, kurtosis, and inter-item polychoric correlations for Subsample 2 are presented in Table 4. The patterns of inter-item polychoric correlations were similar to the patterns displayed in the EFA sample presented in Table 3, except for the inter-item correlations with Item 7. Item 7 showed moderate inter-item correlations with Items 1-6, ranging from  $r = .55$  to  $r = .70$ . Overall, the pattern of correlations did not seem to align clearly with the affective and behavioral aspects or with the “support”, “attachment”, and “experiential learning” categories. Similar to the EFA sample, the “experiential learning” items had low correlations ranging from .184 to .296 among each other. Overall, there was not a clear factor structure based on the polychoric correlations.

An examination of skewness revealed one value that exceeded a recommended value of  $|2|$  (Finney & DiStefano, 2013). Item 4 (“I had at least one professor at [institution] who made me excited about learning”) had a skewness value  $|2.28|$ . An examination of kurtosis revealed no extreme deviations from univariate normality, as all values fell within  $|7|$  (Finney & DiStefano, 2013). Calculations of Mardia’s normalized kurtosis indicated a value of 146.63. This value was greater than recommendations of lower than 10 or 20 (Bentler, 1998), which indicated results of the analyses might be inaccurate if maximum likelihood estimation was used. The evidence of univariate and

multivariate non-normality provided additional support for the use of Robust Diagonally Weighted Least Squares (RDWLS) estimation. RDWLS is more robust than maximum likelihood estimation when data are ordinal and non-normal (Finney & DiStefano, 2013).

Model fit was assessed by examining the  $\chi^2$  statistic, SRMR, RMSEA, and CFI fit indices. Researchers have recommended that fit indices be interpreted as guidelines rather than as strict cutoff values when assessing model fit (Marsh, Hau, & Wen, 2004). The following cutoff values were used as a guideline: SRMR < .08, RMSEA < .06, CFI > .95 (Hu and Bentler, 1998; 1999). In addition to the fit indices, polychoric correlation residuals for pairs of items were examined to assess local fit. If the differences between the observed correlations and the model-implied correlations were large, this indicated the model did not reproduce the correlations well regardless of what of the global fit indices implied (Hu & Bentler, 1999). In the current study, polychoric correlation residuals greater than  $|.15|$  indicated the model did not reproduce the item-pair relationships well.

**Reflective models: Confirmatory factor analysis.** The CFAs for the current study were estimated using Mplus software version 7.3 (Muthén & Muthén, 2014). RDWLS estimation was used to estimate the parameters and fit indices for the model. RDWLS has been found to produce less biased parameter estimates,  $\chi^2$  values, and standard errors than maximum likelihood estimation when data are non-normal and ordinal (Finney & DiStefano, 2013).

Because the alternative models were nested,  $\chi^2$  difference tests were conducted to compare each model. The RDWLS adjusted  $\chi^2$  values were used to compute difference tests. The RDWLS adjustment is similar to that of the Satorra-Bentler adjustment (Finney

& DiStefano, 2013). Because the models were nested within the three-factor model,  $\chi^2$  difference tests were conducted between each alternative model and the three-factor model. If the change in the adjusted  $\chi^2$  was statistically significant, then the more complex model (i.e., the model that estimated more parameters and had fewer degrees of freedom) was considered to fit the data better than the less complex model. As described below, the three-factor model fit better than both the two-factor and one-factor models.

The fit of the three-factor model was supported: adjusted  $\chi^2(32) = 202.81, p < .0001$ ; CFI = .97; RMSEA = .087; SRMR = .05 (Table 8). In addition to the fit indices, the polychoric correlations showed no areas of local misfit. For the three-factor model (Figure 4), there were 55 observations in the polychoric correlation matrix and 23 estimated parameters (10 error variances + 10 path coefficients + 3 factor correlations) resulting in 32 degrees of freedom. All unstandardized and standardized coefficients for the three-factor model were statistically significant and all standardized coefficients were greater than .40, with seven of the ten items having values greater than .70 (Table 9). The structure coefficients indicate the relationships between a particular item and a factor.

The two-factor model consisting of an affective factor and a behavioral factor, fit significantly worse than the three-factor model: adjusted  $\Delta\chi^2(2) = 106.22, p < .0001$ ; CFI = .94; RMSEA = .119; SRMR = .07. The two-factor model had 55 observations in the polychoric correlation matrix and 21 estimated parameters (10 error variances + 10 path coefficients + 1 factor correlation), resulting in 34 degrees of freedom. The polychoric correlation residuals showed 3 areas of local misfit (Table 10). The relationships between Items 1 and 2, 1 and 4, and 1 and 6 showed correlation residuals greater than  $|.15|$  of -.13, -.19, and -.19. The relationships with Item 1 seemed to be overestimated. Item 1 states,

“[Institution] was the perfect school for people like me.” Although Gallup-Purdue (2014) categories this item as “attachment”, it is reasonable to see how this item might relate to some of the “support” items. That is, as was discussed in Chapter 2, it is plausible that a student who feels attached to the institution, also felt supported by faculty and/or mentors while they attended the institution. The areas of local misfit and issues with global misfit indicated that the two-factor model did not reproduce the data well. Consequently, the two-factor model was not supported for the GPI-SEAS.

The one-factor model had 55 observations and 20 estimated parameters (10 error variances + 10 path coefficients), resulting in 35 degrees of freedom. The one-factor model also fit worse than the three-factor model: adjusted  $\Delta\chi^2(3) = 133.58, p < .0001$ ; CFI = .93; RMSEA = .123; SRMR = .08. The polychoric correlation residuals showed 4 areas of local misfit (Table 10). Specifically, the relationships between Items 1 and 2, 1 and 4, 1 and 6, and 8 and 9 showed correlation residuals greater than  $|.15|$ . Similar to the two-factor model, a few relationships between Item 1 and other items were not reproduced well. Like the two-factor model, relationships between Item 1 and Items 2, 4, and 6 were overestimated. The relationship between Items 8 and 9 was underestimated with a residual of .17. The areas of local misfit in conjunction with the fit indices, provided evidence that the one-factor model did not reproduce the data well. This finding is not surprising, given previous research has not shown support for a one-factor structure of student engagement.

Given the good fit of the three-factor model, the pattern coefficients and error terms were examined (Table 9). All unstandardized and standardized pattern coefficients were statistically significant ( $p < .001$ ) and values for standardized coefficients ranged



from .43 to .90. In addition, only three items (Items 8, 9, and 10, the “experiential learning” items) had less than 50% of their variance explained by the model.

Based on CFA results, the three-factor model fit the data well enough to support a three-factor structure. Championing the three-factor model for the GPI-SEAS items indicates support for modeling “attachment”, “support”, and “experiential learning” factors as indicated by Gallup (2014). Reliability estimates were calculated using a three-factor model of the GPI-SEAS. Traditionally, Cronbach’s alpha has been used as a reliability measure for observed composite scores. However, since the items on the GPI-SEAS are complex, it would be misleading to report Cronbach’s alpha because it overestimates reliability when not all systematic variance is due to the latent factor. Internal consistency reliability using McDonald’s (1999)  $\omega$  allows only the variance due to the factor of interest to be treated as systematic variance. For the current study, McDonald’s  $\omega$  was calculated for each of the three factors. The reliability estimates for support, attachment, and experiential learning were all greater than .80 (Table 9). The average proportion of variance in the indicators accounted for (or extracted) by the latent factors was also calculated. When the average proportion of variance extracted is greater than .50, this indicates that the amount of variance measured by the factor is greater than the variance due to measurement error. The variances extracted for attachment, support, and experiential learning were .83, .81, and .52, respectively. The variance measured by each factor was greater than the variance due to measurement error. The correlations among factors ranged from .60 to .77, indicating that the factors were related but somewhat distinct.

**Research question 4.** What is the magnitude and direction of the relationship between student engagement, life satisfaction, and work satisfaction when measured reflectively? Prior to estimating the full structural model, CFA was conducted to assess the structure of the latent constructs. The three-factor CFA model showed adequate overall fit. The polychoric correlation residuals indicate the CFA model reproduces all relationships among the GPI-SEAS items well. The factor correlations, reliability estimates, and variance extracted all provided evidence to support the three-factor CFA model. The three-factor CFA model (Figure 4) was estimated to determine whether a reflective model was appropriate for the GPI-SEAS and to determine the fit of the measurement portion of the theoretical full structural model.

A model with all structural paths estimated (i.e., Model A) was estimated to make to a comparison to the theoretical model. In Benson's (1998) external stage of construct validation, she recommends testing rival hypotheses rather than only one confirmationist model. Model A included the three latent variables examined in the championed CFA model and the life satisfaction and work satisfaction outcome indicators. The observed variable, "life satisfaction", was a single item that asked respondents to rate where they currently stand on a ladder that represents the best or worst possible life for them. Response options range from 0 (*worst possible*) to 10 (*best possible*). The "work satisfaction" variable was an item that stated, "I am deeply interested in the work that I do." Response options ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). The three latent variables were scaled using the metric of the indicators (i.e., attachment was set with indicator 1, support was set with indicator 3, and experiential learning was set with indicator 8). The error terms for the outcome variables were allowed to correlate, as it is

plausible that life satisfaction and work satisfaction are related. Model A, which included relationships specified between each latent variable and each outcome variable can be found in Figure 5.

There were 12 indicators in Model A, thus observations totaled  $12(13)/2 = 78$ . Parameters included 3 correlations among latent variables + 12 error variances + 7 factor loadings + 3 latent variances + 6 structural paths + 1 correlated error term = 32 parameters. Therefore, Model A had 46 degrees of freedom. Table 11 details the fit information for the Model A, adjusted  $\chi^2(46) = 234.75, p < .0001$ ; CFI = .97; RMSEA = .077; SRMR = .047. Although the RMSEA value is slightly above the recommended cutoff value of .06, Model A showed approximate fit. However, two path coefficients for the outcome variables were not statistically significant ( $p > .05$ ). Specifically, the path from the experiential learning factor to the life satisfaction outcome variable ( $\beta = .29, p = .135$ ) and the path from the support factor to the work satisfaction outcome variable ( $\beta = -.02, p = .878$ ) were not statistically significant. Both paths from the support factor to the life satisfaction and work satisfaction outcomes were negative, signaling possible issues with the life satisfaction and work satisfaction outcome variables in the model. All other paths from the latent factors to the outcome variables were statistically significant and standardized coefficients ranged from .18 to .40. Model A only explained 12% ( $R^2 = .12, p < .0001$ ) of the variance in the life satisfaction outcome and 12% ( $R^2 = .12, p < .0001$ ) of the variance in the work satisfaction outcome. Although Model A showed approximate global fit, the GPI-SEAS items did not seem to make major contributions to the outcomes.

Figure 6 illustrates the paths in the theoretical model (i.e., Model B). I hypothesized that attachment and support would influence life satisfaction and support and experiential learning would influence work satisfaction. This hypothesis was based on previous research that showed positive relationships between students who experienced a supportive campus environment and felt the institution was committed to their success and life satisfaction measured one to two years after graduation, although using different instruments related to student engagement (Kuh, 2009; Schmaling & Guy, 2014). Previous research using the Gallup-Purdue Index indicated that students who felt supported by their institution and who participated in experiential learning, were more engaged and satisfied in their jobs than students who did not have those experiences (Ray & Kafka, 2014).

The latent variables were scaled using the same method described for Model A. The degrees of freedom for Model B were calculated using the same number of observations as Model A (i.e., 78 observations) and with the following parameters: 3 correlations among latent variables + 12 error variances + 7 factor loadings + 3 latent variances + 4 structural paths + 1 correlated error term = 30 parameters. Therefore, Model B had 48 degrees of freedom. Model B did not fit significantly worse than the saturated model based on the adjusted  $\chi^2$  difference test, adjusted  $\Delta\chi^2(2) = 5.865, p = .053$  (Table 11). Model B showed approximate fit similar to that of Model A: CFI = .97; RMSEA = .075; SRMR = .048. In addition to the global fit indices, there were no polychoric correlation residuals greater than  $|.15|$ , indicating the model showed good local fit. However, the path from support to work satisfaction was not statistically significant ( $p = .170$ ), which failed to support my hypothesis. All other path coefficients

from the latent factors to the outcomes variables were statistically significant and standardized coefficients ranged from .12 to .43 (Table 12). Model B only explained 11% ( $R^2 = .11, p < .0001$ ) of the variance in the life satisfaction outcome and 13% ( $R^2 = .13, p < .0001$ ) in the work satisfaction outcome.

### **Stage III: Estimating Formative Models**

**Research question 5.** How well does a formative model represent the GPI-SEAS items? Answering research question 5 required estimating three formative models treating the GPI-SEAS items as formative indicators. The life satisfaction and work satisfaction items were treated as effect indicators in each model, with paths included from each latent variable to each effect indicator. Models with formative indicators and effect indicators as described, are typically referred to as multiple indicator-multiple cause (MIMIC) models (Jöreskog & Goldberger, 1975). The three models that were tested included: 1) a model with one formative construct representing general student engagement; 2) a model with two formative constructs representing affective and behavioral aspects of engagement; and 3) a model with three formative constructs representing support, attachment, and experiential learning.

The MIMIC models were estimated using Mplus software version 7.7 (Muthén & Muthén, 2014). Like the reflective models, RDWLS estimation was used to estimate the parameters and fit indices for the formative models using Subsample 2. Model fit was assessed by examining the RDWLS adjusted  $\chi^2$  statistic, RMSEA, and CFI fit indices. The following cutoff values were used as a guideline to assess model fit: RMSEA  $< .06$ ; CFI  $> .95$  (Hu & Bentler, 1998; 1999). In addition to the fit indices, I examined the path coefficients, the variance explained (using overall  $R^2$ ) for each formative construct, and

multicollinearity (Bollen, 2011). Standardized and unstandardized coefficients were examined to determine whether they were statistically significant (Bollen, 2011). The  $R^2$  values for the formative constructs were examined to determine the amount of variance explained by the formative indicators. Multicollinearity among the indicators was also examined. Ideally, the indicators should have low collinearity, otherwise it is difficult to estimate their individual effects (Bollen, 2011).

Overall, the model with one formative construct showed good global fit. However, the models with two- and three-constructs did not fit well globally. For the model with three-constructs, there were 78 observations with 37 estimated parameters (2 error terms + 3 disturbance terms [each fixed to 0] + 6 total factor pattern coefficients [3 fixed to 1.0] + 10 directional paths + 10 exogenous variances + 12 exogenous covariances), resulting in 41 degrees of freedom. The three formative constructs were scaled using the metric of the outcome variables (i.e., attachment, support, and experiential learning were each set with the work satisfaction outcome variable). The attachment construct consisted of four items and both the support and experiential learning constructs each had three items. The formative indicators for each formative construct were allowed to freely covary with one another, but not across formative constructs. For example, the formative indicators for attachment (i.e., Items 1, 2, 5, and 7) could covary with one another, but could not covary with the support or experiential learning formative indicators (i.e., Items 3, 4, and 6 and Items 8-10, respectively). The three disturbance terms were fixed to 0, which essentially assumes that there is no measurement error in the model. An examination of the results for the three-construct

model indicated poor model fit: adjusted  $\chi^2(41) = 868.65$ ; CFI = .43; RMSEA = .17 (see Table 13 for all formative model fit indices).

The two-construct model had 78 observations with 48 estimated parameters (2 error terms + 2 disturbance terms for the latent constructs [each fixed to 0] + 12 factor pattern coefficients + 10 exogenous variances + 24 exogenous covariances), resulting in 30 degrees of freedom. The two constructs were scaled using the metric of the outcome variables (i.e., the affective and behavioral constructs were each set with the work satisfaction outcome variable). The affective construct included seven items (i.e., the combined support and attachment items) and the behavioral construct included three items (i.e., the experiential learning items in the three-construct model). The formative indicators were allowed to freely covary with one another, but not across constructs. The two disturbance terms were set to zero, similar to the three-construct model. An examination of results for the two-construct model indicated poor model fit: adjusted  $\chi^2(30) = 328.56$ ; CFI = .79; RMSEA = .12.

The two- and three-construct models with covariances constrained for indicators between constructs provided additional degrees of freedom for each model. However, constraining the covariances between construct indicators to zero probably does not make theoretical sense because the polychoric correlation matrix (Table 3) showed statistically significant relationships among indicators across constructs as discussed earlier in the chapter. Therefore, the two- and three-construct models were estimated allowing all formative indicators to covary. The two- and three-latent construct models could only be estimated in Mplus when the disturbance terms for each construct (i.e., the behavioral and affective constructs in the two-construct model and the support, attachment, and

experiential learning constructs in the three-construct model) were set to zero. But, setting the disturbance terms to zero resulted in a lack of necessary  $R^2$  values for the formative constructs in each model. To obtain  $R^2$  values for each formative construct, an additional path from each respective formative construct to the life satisfaction outcome was set to one. For example, in the two-construct model, to obtain the  $R^2$  value for the affective construct, the disturbance term was unconstrained, but the path from the affective construct to the life satisfaction outcome was set to one. Then, the same method of freeing/constraining parameters was used for the behavioral construct to obtain the  $R^2$  value for that construct.

Although constraining additional parameters allowed me to obtain  $R^2$  values, freeing/constraining various parameters changed the model interpretations. Therefore, although the two- and three-construct models showed acceptable global fit, they were rejected due to issues with model estimation and interpretation. Thus, the one-construct model with freely covarying formative indicators was championed as the best fitting formative model. Fit indices for all formative models are reported in Table 13.

The two-construct model (Figure 8) had 78 observations with 69 estimated parameters (2 error variances + 2 factor pattern coefficients + 10 directional paths + 10 exogenous variances + 45 exogenous covariances), resulting in 9 degrees of freedom. The two formative constructs were scaled in the same way as the previous two-construct model (i.e., the two latent variables were set with the work satisfaction effect indicator). The two-construct model showed approximate global fit: adjusted  $\chi^2(9) = 79.03$ ; CFI = .95; RMSEA = .11 (Table 13). The RMSEA fit index did not meet recommended values, but the CFI fit index indicated approximate fit. The standardized coefficients for Items 4



(“I had at least one professor at [institution] who made me excited about learning”) and 9 (“While attending [institution], I worked on a project that took a semester or more to complete”) were not statistically significant, at  $p < .05$ . There were also four negative standardized coefficients associated with Items 2 ( $\beta = -3.05, p < .05$ ), 3 ( $\beta = -.59, p < .05$ ), 4 ( $\beta = -.02, p = .84$ ), and 5 ( $\beta = -.93, p < .05$ ). The behavioral/experiential learning construct was not a statistically significant predictor of life satisfaction ( $\beta = .01, p = .91$ ) or work satisfaction ( $\beta = .13, p = .05$ ). The seven formative indicators (Items 1-7) specified for the affective construct explained 21% ( $R^2 = .21, p < .0001$ ) of the variance in the construct. The three formative indicators (Items 8-10) specified for the behavioral construct did not explain a statistically significant amount of variance,  $R^2 = .03, p = .14$ .

The three-construct model (Figure 9) had 78 observations with 70 estimated parameters (2 error variances + 3 factor pattern coefficients + 10 directional paths + 10 exogenous variances + 45 exogenous covariances), resulting in 8 degrees of freedom. The three formative constructs were scaled in the same way as the previous three-construct model (i.e., the three latent variables were set with the work satisfaction indicator). The three-construct model showed similar fit to the two-construct model with all indicators allowed to freely covary: adjusted  $\chi^2(8) = 84.40$ ; CFI = .95; RMSEA = .12 (Table 13). The RMSEA fit index did not meet the recommended value. Three standardized coefficients that were not statistically significant. Specifically, none of the formative indicators for the experiential learning construct (Items 8, 9, and 10) were statistically significant, and the standardized coefficient for Item 10 (“I was extremely active in extracurricular activities and organizations while attending [institution]”) was negative (i.e.,  $\beta = -0.05, p = .93$ ). Items 1, 2, 3, and 7 also had negative standardized coefficients.

Additionally, the experiential learning latent variable was not a statistically significant predictor of the life satisfaction ( $\beta = -0.07, p = .24$ ) or work satisfaction ( $\beta = .02, p = .72$ ) outcome variables. Four formative indicators (Items 1, 2, 5, and 7) explained 22% ( $R^2 = .22, p < .0001$ ) of the variance in the attachment indicator. Three formative indicators (Items 3, 4, and 6) explained a non-significant amount of the variance in the support construct ( $R^2 = .03, p = .14$ ). Finally, three formative indicators combined (Items 8-10) explained a statistically non-significant amount of the variance in the experiential learning construct ( $R^2 = .02, p = .17$ ).

As a reminder, although the two- and three-construct models showed approximate global fit, they were rejected due to issues with estimation and clear model interpretation. Thus, the one-construct model was championed as the best-fitting formative model. The one-construct model had 78 observations with 69 estimated parameters (2 error variances + 1 disturbance term + 1 factor pattern coefficient + 10 directional paths + 10 exogenous variances + 45 exogenous covariances), resulting in 9 degrees of freedom. The one-construct model was scaled similarly to the two- and three-construct models, using the metric of the outcome variables (i.e., the engagement construct was set with the work satisfaction outcome variable). All 10 formative indicators were allowed to freely covary in the one-construct model. The fit indices indicated good global fit: adjusted  $\chi^2(9) = 24.72$ ; CFI = .99; RMSEA = .05.

Although the global fit for the one-construct model was good, six of the ten unstandardized and standardized coefficients were non-significant at  $p < .05$ . (Table 14). Formative indicators 2, 3, 4, 7, 9, and 10 showed non-significant standardized coefficients. The items with non-significant coefficients included two “attachment”

items, two “support” items, and two “experiential learning” items. Additionally, the standardized coefficient values for indicators 2, 3, and 4 were negative at  $-.10$  ( $p = .117$ ),  $-.12$  ( $p = .096$ ), and  $-.05$  ( $p = .369$ ), respectively. The negative coefficients indicate the model is overestimating the relationships for these items with the general engagement construct. Recall that the coefficients for items 2, 3, and 4 were also non-significant. Item 2 stated “I can’t imagine a world without [institution]” and was categorized by Gallup-Purdue (2014) as “attachment”. Item 3 stated “My professors at [institution] cared about me as a person” and was categorized as support”. Item 4 stated “I had at least one professor at [institution] who made me excited about learning” and was categorized as “support”. The 10 formative indicators explained only 29% of the variance in the general engagement construct ( $R^2 .29, p < .0001$ ). With that said, the engagement latent variable was a statistically significant predictor of both life satisfaction ( $\beta = .56, p < .001$ ) and work satisfaction ( $\beta = .60, p < .001$ ). However, the negative and non-significant standardized coefficients may point to possible issues with multicollinearity and provide evidence that there may be issues with the indicators included in the one-construct formative in the model.

I used a guideline of polychoric correlation values  $>.90$  to indicate multicollinearity (Tabachnick & Fidell, 2013). An examination of the polychoric correlation matrix for Subsample 2 (Table 4) indicated that none of the values were multicollinear. However, the polychoric correlation for Item 1 (“[Institution] was the perfect school for people like me”) and Item 2 (“I can’t imagine a world without [institution]”) was  $.730, p < .001$ . The polychoric correlation for Item 3 (“My professors at [institution] cared about me as a person”) and Item 4 (“I had at least one professor at

[institution] who made me excited about learning”) was .689,  $p < .001$ . Although Items 2, 3, and 4 were not redundant according to the cutoff value, the polychoric correlations were moderate and may be an explanation for the issues seen with these items.

**Research question 6.** What is the magnitude and direction of the relationship between student engagement, life satisfaction, and work satisfaction when measured formatively? To address research question 6, I examined standardized paths and the  $R^2$  (Table 14) values for the one-latent variable model. The standardized coefficients were statistically significant for both life satisfaction ( $\beta = .56, p < .001$ ) and work satisfaction ( $\beta = .60, p < .001$ ). The paths indicated that for every one standardized unit increase in engagement, both life satisfaction and work satisfaction are expected to increase by approximately .6 standardized units. In addition, the one-construct model explained 31% ( $R^2 = .31, p < .0001$ ) of the variance in life satisfaction and 36% ( $R^2 = .36, p < .0001$ ) of the variance in work satisfaction. The magnitude of the path coefficients and the  $R^2$  values for the outcome variables provided some evidence that the GPI-SEAS items predict the intended outcomes of life and work satisfaction.

## CHAPTER 5

### Discussion

Student engagement is a complex construct that is thought to be related to positive outcomes during and after college. In the current research study, I focused on the measurement of alumni's perceptions of indicators of their engagement during college and implications of their responses on the validity of scores for the GPI-SEAS. Consisting of three stages, the current research study included an evaluation of two competing models that could be used to conceptualize and measure engagement and an examination of validity evidence for the GPI-SEAS items. In Stage I, the inter-item correlations and EFA showed support for a one-factor solution for the GPI-SEAS items. However, after comparisons of nested reflective CFA models in Stage II, a three-factor model was championed. In Stage III, formative MIMIC models were estimated and the results supported a one-latent construct model. The results across each of the three stages conflicted with one another and explanations for each stage are included in subsequent sections.

The discussion of the results is organized in the following way. First, the results of the research questions subsumed under Stage I are discussed along with implications for validity evidence of the GPI-SEAS scores. Second, the results of the research questions subsumed under Stage II are discussed and conclusions regarding the use of a reflective model for the scale are made. Third, the results of the research questions subsumed under Stage III are discussed along with conclusions about the use of a formative model for the scale. In addition, the results from the latter two stages are used to make recommendations to higher education researchers and practitioners interested in

measuring student engagement and making connections to post-college outcomes. Finally, limitations of the current study and suggestions for future research are outlined.

### **Stage I**

Stage I included an analysis of the GPI-SEAS items and an exploratory factor analysis on a subsample of the respondents. The two research questions included in Stage I pertained to 1) the pattern of inter-item correlations for the GPI-SEAS and 2) the number and nature (i.e., pattern of loadings) of the factors that accounted for the inter-item correlations.

The polychoric correlations did not show clear evidence for a three-factor model (based on Gallup-Purdue's attachment, support, and experiential learning categories) or a two-factor model (based on behavioral and affective aspects supported by previous research). That is, the three items (Items 8, 9, and 10) expected to form the experiential learning/behavioral factors had low inter-item polychoric correlations (i.e.,  $r = .182$  to  $r = .311$ ,  $p < .05$ ), suggesting those items might not group together as a factor in subsequent analyses. The content of Items 8, 9, and 10 included having an "... internship or job that allowed me to apply what I was learning in the classroom", "... work[ing] on a project that took a semester or more to complete", and being "... extremely active in extracurricular activities and organizations", respectively. Based on the range of types of experiences included in the three items, it is plausible that respondents who experienced one activity may not have experienced the others, consequently leading to low inter-item correlations.

In addition to the low correlations among the "experiential learning" items, several of the "support" and "attachment" items showed moderate inter-item correlations.

Although the moderate correlations between the “support” and “attachment” items provided some evidence to combine them into an “affective” factor, the correlations did not provide clear support for items as separate “support” and “attachment” factors. Overall, the inter-item polychoric correlations seemed to suggest one-factor, which was subsequently supported with exploratory factor analyses.

Due to the lack of validity evidence and psychometric investigations of the GPI-SEAS, exploratory factor analytic techniques were used to examine the internal structure of the instrument. Recall that EFA is considered reflective, because changes in the latent construct (i.e., student engagement) are thought to precede changes in the observed variables. Several EFA solutions were examined based on the results of the parallel analysis and Gallup-Purdue’s (2014) conceptualization of the scale. Specifically, one-, two-, and three-factor solutions were examined using interpretability as the primary basis for choosing among the models. Of the three solutions, the one-factor model was considered the most interpretable.

For the results to provide supporting structural validity evidence for the GPI-SEAS (Benson, 1998), the factors emerging from the data needed to align with one of the hypothesized groupings of the GPI-SEAS items. That is, the factors needed to align with the two interrelated aspects of engagement (i.e., affective and behavioral) described in the student engagement literature (Butler, 2011; Chapman, 2003; Handelsman et al., 2005; Kuh, 2003; Mandernach, 2015). Recall the behavioral and affective aspects of engagement are related to students’ participation in educationally effective activities and students’ perceptions of features of the institutional environment that support their

development, respectively (Kuh, 2009; McCormick, Kinzie, & Gonyea, 2013; Harper & Quaye, 2015).

If the factors did not support the theoretical behavioral and affective aspects of engagement, they needed to align with the three areas (i.e., support, attachment, and experiential learning) labeled by Gallup-Purdue (2014). Unfortunately, none of the EFA results including the one-factor solution aligned with the conceptualizations put forth by previous literature. The equivocal polychoric correlations provided evidence as to why the EFA did not result in a two- or three-factor solution. The moderate inter-item correlations suggested that there may not be two or three distinct dimensions of student engagement for the GPI-SEAS items. To further test the structure of the GPI-SEAS, confirmatory factor analyses were conducted.

## **Stage II**

Stage II included a comparison of nested CFA models on a second subsample of the respondents. The two research questions subsumed under Stage II pertained to 1) the dimensionality of the GPI-SEAS items using reflective models and 2) the magnitude and direction of the relationship between student engagement and the outcomes, life satisfaction and work satisfaction, when measured reflectively.

A series of nested CFA models were estimated and the three-factor model fit statistically significantly better than one- and two-factor models. The three-factor model showed approximate global and local fit and seemed to adequately represent the relationships among the GPI-SEAS items. The model also provided support for Gallup-Purdue's distinction between the three areas of attachment, support, and experiential



learning. That is, the responses to the items can be meaningfully represented with three separate scores based on each factor.

The reflective CFA model of the GPI-SEAS provides support for Benson's (1998) structural stage of construct validity. Given a three-factor structure, I sought to provide evidence for Benson's external stage by examining the relationship between GPI-SEAS scores, life satisfaction and work satisfaction. The hypotheses regarding how the GPI-SEAS should relate to the outcome variables were not fully supported. Specifically, the support factor had a non-significant relationship with work satisfaction and a negative relationship with life satisfaction. Support should be related to work satisfaction, but this finding was not supported by the current study (Gallup-Purdue, 2014; Ray & Kafka, 2014). The findings could be a function of the sample used in the current study. Perhaps, for this sample, "support" is not related to work satisfaction. Or, the current findings could be related to the specific item that was chosen to represent "work satisfaction". The work satisfaction outcome item stated, "I am deeply interested in the work that I do". It is possible that this particular item is not connected to "support," but other items related to work satisfaction are related to support. Including additional observed variables related to work satisfaction in future studies, would provide additional evidence as to whether the outcome is related to support.

Further, the reflective model explained 11% of the variance in the life satisfaction outcome and 13% in the work satisfaction outcome. Although the model does not seem to explain a sizable percentage in either outcome, the amount of variance explained in each outcome variable indicates some utility of the model.

### Stage III

Stage III included a comparison of formative MIMIC models on the second subsample of respondents. The two research questions in Stage III pertained to 1) how well a formative model fit the GPI-SEAS items and 2) the magnitude and direction of the relationship between student engagement, life satisfaction, and work satisfaction when measured formatively.

One-, two-, and three-construct formative models were estimated with the GPI-SEAS items as formative indicators and life satisfaction and work satisfaction as outcome variables. The one-construct model was championed as having the best fit. However, several of the formative indicators had either negative or non-significant standardized and unstandardized coefficients. The negative and non-significant coefficients for the formative indicators point to possible issues when the GPI-SEAS items are modeled formatively. Items 2, 3, and 4 had both negative and non-significant coefficients, indicating that these items should be evaluated more closely. Item 2 was categorized as attachment and stated, “I can’t imagine a world without [institution]”. Item 3 was categorized as support and stated, “My professors at [institution] cared about me as a person”. Item 4 was also categorized as support and stated, “I had at least one professor at [institution] who made me excited about learning”. It is not completely clear why Items 2, 3, and 4 had issues in the model. However, the results may suggest that these items should be measured reflectively rather than formatively. That is, Items 2, 3, and 4 showed moderate inter-item correlations ( $r_{2,3} = .49$ ,  $r_{2,4} = .36$ , and  $r_{3,4} = .69$ , respectively). In the formative model, all items were allowed to correlate, which may explain some of the issues with the model. Further, the standardized paths in the three-factor reflective model

for Items 2, 3, and 4 ( $\beta = .74$ ,  $\beta = .89$ , and  $\beta = .76$ , respectively) were much stronger than the one-construct formative paths.

Although there are issues with the one-construct model, the model showed statistically significant relationships between the latent construct and both life satisfaction and work satisfaction, explaining 31% and 36% of the variance in each respective outcome. A statistical test was not used to compare the amount of variance explained in the outcomes for the reflective and formative models. However, the formative model explained a greater proportion of the variance in both the life satisfaction and work satisfaction outcomes. With that said, the negative and non-significant coefficients for the formative indicators provide evidence that the GPI-SEAS may not be measured most effectively as a formative construct.

### **Recommendations for the Use of the GPI-SEAS**

The reflective and formative models provided divergent results. The reflective approach provided support for a three-factor model of the GPI-SEAS items, but the formative approach provided support for a one-latent construct model. Depending on which method someone chooses, the results will differ. Thus, researchers who choose one approach over the other, would likely come to different conclusions about the GPI-SEAS items. For example, someone who chooses to model the GPI-SEAS items formatively, might determine that the items do not function well based on the model and perhaps some items should be removed prior to additional analyses or from the instrument entirely. However, someone who models the GPI-SEAS items reflectively, might conclude that the items generally seem fine. The opposing conclusions based on the type of approach used, are a cause for concern from an empirical standpoint. Thus, it is necessary to also

evaluate the GPI-SEAS items from a theoretical perspective using the Jarvis et al. (2003) criteria discussed in chapter two.

Recall that based on Jarvis et al.'s (2003) criteria for choosing between formative and reflective models, from a theoretical sense, some of the GPI-SEAS items should be measured formatively and some reflectively. Or, the GPI-SEAS items should be modeled with multiple dimensions. The latter explanation fits the use of the three-factor reflective model. Additionally, there were fewer issues with the unstandardized and standardized coefficients for the reflective model than for the formative model. However, when modeled reflectively a smaller percentage of the variance was explained for the life satisfaction and work satisfaction outcomes. I would consider it premature to completely rule out the formative model, but for the sample used in the current study, the reflective model fit better empirically and makes more theoretical sense. Thus, it seems appropriate to compute subscale scores for each of the three factors: attachment, support, and experiential learning. Computing separate scores for each dimension reflects the multidimensionality of the items and their potentially different relationships with other variables.

The three factors of attachment, support, and experiential learning capture important aspects of student engagement, such as participation in educationally effective activities and students' perceptions of aspects of the institutional environment that support their learning (Butler, 2011; Chapman, 2003; Handelsman, et al., 2005; Kuh, 2003; Mandernach, 2015). Further, the GPI-SEAS items seem to make an important distinction between having meaningful relationships with faculty or mentors (i.e., support) and feeling that the institution cares about students (i.e., attachment). While

support and attachment are related, this study provides evidence that the areas are distinct, and shouldn't be combined to represent an affective or perceptual domain.

### **Limitations of the Current Study and Directions for Future Research**

There were several limitations of the current study that could not be overcome due to sample availability and other factors. First, the data for the GPI-SEAS were collected from alumni respondents who were asked to reflect on their experiences from when they were in college. Respondents graduated from the institution between 10 and 20 years prior to the administration of the survey. It is possible that respondents' memories were imprecise or that their responses were influenced by their current state in life rather than actual experiences while they were at the institution. Thus, it is plausible that some responses may not accurately reflect respondents' experiences. Data on similar instruments (e.g., the NSSE, CSEQ) from the same sample while they attended college would allow for comparisons to be made between their responses from different points in time. Although the instruments would not be the same, if respondents showed similarities in their answers to items from different occasions, additional validity evidence could be provided for the GPI-SEAS scores.

Second, graduates were asked to reflect on their undergraduate experiences, but they did so through the lens of their current lives. Therefore, graduates' current experiences helped to shape how they viewed their past experiences. We assume that graduates' past experiences during college influence how they perceive their life satisfaction and work satisfaction. However, their current life satisfaction and work satisfaction likely influences how they view their past experiences. Thus, the relationship between graduates' perceptions of their past experiences and their current life

experiences is potentially recursive. In an ideal methodological framework, data should be collected at two separate time points. Specifically, data on student experiences should be collected while students are still at the institution. Inferences made from responses while students attended the institution would be more trustworthy than inferences made from responses from graduates reflecting back 10 to 20 years after graduation. If researchers were also interested in connecting student experiences to post-college outcomes, a longitudinal design would be necessary. Designing a longitudinal study would take more planning at the early stages of the research study, but it would allow researchers to match samples from multiple time points and make more trustworthy inferences related to student engagement during college and its relation to post-college outcomes.

Third, the sample consisted of predominantly White respondents. It is possible that the results would have been different if a more racially or ethnically diverse sample was used. For example, race/ethnicity has been shown to moderate the relationship between student engagement and career earnings (Hu & Wolniak 2013). Specifically, engagement had a positive and statistically significant relationship on earnings for Native American and Latino students, but no relationship among African American or Asian American students (Hu & Wolniak, 2013). It is thus conceivable that student engagement might have a different relationship with other outcomes related to student engagement based on race/ethnicity. Future studies should conduct analyses using more diverse samples. Confidence in the results of the current study would be increased if similar findings emerged in other formative or reflective analyses using more racially/ethnically diverse samples.

Fourth, on the current GPI instrument, respondents see both engagement and outcome-related questions on the same survey. As mentioned previously, respondents' current state in life likely influences their responses to both the engagement and outcome items, which threatens inferences made from the responses. If the relationship between respondents' current lives and their reflection on past experiences is recursive, then establishing temporal precedence is problematic. Specifically, I cannot say with certainty whether respondents' engagement during college influences their responses to the items on the survey or whether respondents' current life conditions influence their responses to the items. Thus, because responses to the outcome items were collected at the same time as responses to the engagement items (asking respondents to reflect on past experiences), it may be misleading to consider the life satisfaction and work satisfaction variables as true outcomes. Again, future research should consider a longitudinal research design in which responses to engagement items are collected while students are at the institution and responses to post-college outcome items are collected after students have graduated from the institution. A longitudinal design would not only allow us to have more confidence in the inferences made from the responses, but it would also allow researchers to view the life satisfaction and work satisfaction variables as true outcomes.

Fifth, there is currently no published information about the inclusion of the GPI-SEAS items. That is, it is unclear how the items originated, who created them, or why they were deemed important to include on the instrument. Information related to the theory and background of the items is crucial to Benson's (1998) substantive stage of construct validation. Future research provided by Gallup and Purdue University should include information on item construction and the theoretical basis for the inclusion of

specific items on the instrument. Providing background information on the items will contribute additional validity evidence for the GPI-SEAS and will allow researchers to test whether future analyses of the GPI-SEAS items align with the theory or rationale used to create the instrument.

Sixth, no items were removed from the instrument due to redundancy or low inter-item correlations for the analyses. In future studies, researchers should consider removing items that may distract from the scale's purpose. Removing items should be based on theoretical or empirical support such as inter-item correlations that show redundancy or minimal to no relationship with other items. When altering an instrument by removing items, it is important to be aware of the balance between construct-irrelevance (i.e., removing items that distract from the instrument's purpose) and construct underrepresentation (i.e., removing so many items that full coverage of the construct is lost; AERA, APA, NCME, 2014). If items are removed from the GPI-SEAS that are thought to be distracting or unnecessary, writing additional items or revising current items to cover the breadth of the construct may thus be warranted.

Seventh, the use of formative models for measuring constructs such as student engagement, requires the use of outcome variables to identify the model (i.e., the number of observations must be greater than the number of estimated parameters, and each latent variable must have at least two separate indicators). In the current study, life satisfaction and work satisfaction were the outcome variables used to identify or form the formative GPI-SEAS model. If different outcome variables were selected, the overall model fit and interpretation may have changed entirely. For example, student engagement has been shown to have a differential relationship with post-graduate career earnings depending on



gender, race/ethnicity, and SAT/ACT scores (Hu & Wolniak, 2013). If career earnings were used as an outcome variable instead of life satisfaction or work satisfaction, the relationship between earnings and the latent variables might have been positive, negative, or non-significant depending on the demographics of the sample used in the study.

This limitation of the formative approach means that constructs modeled using formative measurement may only be useful insofar as the outcome variables used to identify the model are of interest to the researcher. Therefore, the results of constructs measured formatively may be less valuable than those of constructs measured in a more generalizable way, such as with reflective measurement.

Eighth, the sample of respondents in the current study did not always use the full response scale for all 10 items. Specifically, for some items, the lower scale categories (i.e., 1 and 2) only had a few responses. It is not necessarily a problem that responses were skewed toward the upper end of the response scale for some items. However, it is possible that there are self-selection criteria that led to skewed results for a few items. Recall that the response rate prior to deletion of missing cases was 12.8% and 10% after deletion. Respondents who were the most engaged might have self-selected to respond to the survey, rather than alumni who were not engaged. Future studies should consider additional demographic information (e.g., major, GPA, gender) of those who responded to the survey to see whether the sample reflects the overall institutional demographics. It is conceivable that the sample primarily included respondents who had positive experiences at the institution, and those respondents may or may not represent the total population. Additionally, future research should closely evaluate alumni who responded towards the lower end of the scale across items. An evaluation of respondents who

answered the questions negatively may provide additional insight into what experiences and behaviors related to engagement are most (or least) beneficial to current students and alumni.

### **Implications and Conclusions**

The GPI-SEAS is still in its early stages in terms of empirical evaluation, but it shows potential for measuring student engagement. The GPI-SEAS has evidence from the reflective model aligned with Benson's (1998) structural stage of construct validation to support Gallup-Purdue's conceptualization of the areas of attachment, support, and experiential learning. The current study contributed initial evidence of validity aligned with Benson's (1998) structural and external stages. This dissertation also provides evidence to support using reflective models for the GPI-SEAS items. Although the current study provides support for reflective models with the GPI-SEAS, that does not indicate reflective models are appropriate for every instrument purported to measure student engagement. Researchers and practitioners must be cognizant of the definitions they use for constructs *prior* to determining how to measure them. The criteria outlined by Jarvis et al. (2003) and relevant theory should be used when choosing between formative and reflective models to measure student engagement.

Student engagement instruments are widely used and there is interest among student affairs practitioners and university administrators to bolster student engagement initiatives at colleges and universities. Because the GPI-SEAS can be linked to positive post-college outcomes, the instrument may be used as an additional resource for institutions to gauge which experiences and behaviors should be emphasized among their populations of students. Furthermore, use of the GPI-SEAS may help practitioners target

interventions for current students towards activities and experiences that may have an impact on their engagement. The current research study provided a foundation for the exploration of the GPI-SEAS items and the role of student engagement in post-college outcomes.

Table 1

*A Framework for Assessing Reflective and Formative Models*

Considerations	Reflective Model	Formative Model	Relevant Literature
Nature of Construct	Latent construct is existing <ul style="list-style-type: none"> <li>Latent construct exists independent of the measures used</li> </ul>	Latent construct is formed <ul style="list-style-type: none"> <li>Latent construct is determined as a combination of its indicators</li> </ul>	Borsboom et al. (2003)
Direction of causality between items and latent construct	Causality from construct to items <ul style="list-style-type: none"> <li>Variation in the construct causes variation in the item measures</li> </ul>	Causality from items to construct <ul style="list-style-type: none"> <li>Variation in the construct does not cause variation in the item measures</li> </ul>	Bollen & Lennox (1991); Edwards & Bagozzi (2000); Jarvis et al. (2003)
Characteristics of items used to measure the construct	Items are manifested by the construct <ul style="list-style-type: none"> <li>Items share a common theme</li> <li>Items are interchangeable</li> </ul>	Items define the construct <ul style="list-style-type: none"> <li>Items need not share a common theme</li> <li>Items are not interchangeable</li> </ul>	Jarvis et al. (2003)
Item intercorrelation	Items should have high positive intercorrelations  Empirical test: internal consistency and reliability assessed via coefficient alpha, average variance extracted, and factor loadings	Items can have any pattern of intercorrelations  Empirical test: indicator reliability cannot be assessed empirically; various preliminary analyses useful to check directionality between items and construct	Diamantopoulos & Siguaw (2006)
Item relationships with construct antecedents and consequences	Items have similar sign and significance of relationships with the antecedents/consequences as the construct  Empirical test: content validity is established based on theoretical considerations and assessed empirically	Items may not have similar significance of relationships with the antecedents/consequences as the construct  Empirical test: validity can be assessed empirically using a MIMIC model, and/or structural linkage with another criterion variable	Diamantopoulos (2006)

*Note:* Table adapted from Coltman et al. (2008).

Table 2

*Alignment of the GPI-SEAS Items with Gallup-Purdue Categories and Theoretical Affective and Behavioral Categories*

Items	Gallup-Purdue Category	Affective or Behavioral Category
1. [Institution] was the perfect school for people like me	Attachment	Affective
2. I can't imagine a world without [institution]	Attachment	Affective
3. My professors at [institution] cared about me as a person	Support	Affective
4. I had at least on professor at [institution] who made me excited about learning	Support	Affective
5. [Institution] prepared me well for life outside of college	Attachment	Affective
6. While attending [institution], I had a mentor who encouraged me to pursue my goals and dreams	Support	Affective
7. [Institution] is passionate about the long-term success of its students	Attachment	Affective
8. While attending [institution], I had an internship or job that allowed me to apply what I was learning in the classroom	Experiential Learning	Behavioral
9. While attending [institution], I worked on a project that took a semester or more to complete	Experiential Learning	Behavioral
10. I was extremely active in extracurricular activities and organizations while attending [institution]	Experiential Learning	Behavioral

Table 3

*Item Polychoric Correlations and Descriptive Statistics for Subsample 1 (N = 349)*

Items	1	2	3	4	5	6	7	8	9	10
	A	A	S	S	A	S	A	EL	EL	EL
1	-									
2	.722	-								
3	.569	.488	-							
4	.494	.474	.603	-						
5	.452	.395	.544	.600	-					
6	.352	.309	.430	.440	.586	-				
7	.230	.258	.379	.465	.511	.471	-			
8	.311	.232	.352	.323	.342	.250	.286	-		
9	.111	.125	.150	.177	.210	.242	.260	.311	-	
10	.368	.352	.297	.262	.246	.298	.259	.224	.182	-
M	4.41	3.95	3.90	4.58	4.17	3.15	3.94	3.45	3.48	3.71
SD	.85	1.22	.95	.71	.91	1.29	.95	1.55	1.60	1.22
Skewness	-1.52	-0.97	-1.10	-0.69	-0.67	-2.14	-0.07	-0.50	-0.51	-0.62
Kurtosis	1.93	-0.16	0.94	0.17	0.06	5.92	-1.04	-1.30	-1.36	-0.62

Note: A = Attachment item; S = Support item; EL = Experiential Learning item. All polychoric correlations are statistically significant at  $p < .05$

Table 4

*Item Polychoric Correlations and Descriptive Statistics for Subsample 2 (N = 700)*

Items	1	2	3	4	5	6	7	8	9	10
	A	A	S	S	A	S	A	EL	EL	EL
1	-									
2	.730	-								
3	.493	.490	-							
4	.361	.362	.689	-						
5	.638	.531	.605	.533	-					
6	.327	.343	.641	.574	.455	-				
7	.637	.609	.652	.553	.702	.545	-			
8	.176	.147	.263	.274	.332	.332	.295	-		
9	.163	.147	.276	.275	.281	.236	.283	.296	-	
10	.242	.219	.202	.169	.248	.278	.221	.207	.184	-
M	4.36	3.95	3.93	4.62	4.21	3.11	3.95	3.33	3.64	3.65
SD	.83	1.16	.92	.69	.86	1.37	.94	1.55	1.51	1.29
Skewness	-1.31	-0.98	-0.64	-2.28	-1.02	-0.10	-0.68	-0.36	-0.68	-0.56
Kurtosis	1.37	0.08	0.03	6.17	0.72	-1.19	-0.02	-1.40	-1.07	-0.87

Note: A = Attachment item; S = Support item; EL = Experiential Learning item. All polychoric correlations were statistically significant at  $p < .001$ .

Table 5

*Summary of Results across All Stages of Analyses*

Stage of Analysis	Analysis Performed	Sample	General Results	Additional Details
Stage I	Inter-item correlations	Subsample 1	No clear pattern of inter-item correlations	--
Stage I	1-, 2-, and 3-factor EFAs	Subsample 1	1-factor solution retained	2- and 3-factor solutions were not interpretable
Stage II	Nested 1-, 2-, and 3-factor CFAs	Subsample 2	3-factor model championed	--
Stage II	Full Structural Models with 3-Factor CFA	Subsample 2	Theoretical model showed approximate fit	Validity evidence provided for the 3-factor CFA model
Stage III	1-Construct Formative Model	Subsample 2	1-construct model championed	Items 2, 3, and 4 had negative and non-significant path coefficients
Stage III	2-Construct Formative Model	Subsample 2	RMSEA did not meet recommended values, and several indicators with negative and/or non-significant paths	Non-significant (Items 4 & 9) and negative (Items 2-5) path coefficients
Stage III	3-Construct Formative Model	Subsample 2	RMSEA did not meet recommended values, and several indicators with negative and/or non-significant paths	Non-significant (Items 8-10) and negative (Items 1-3, 7, & 10) path coefficients

*Note.* Subsample 1 included 349 respondents and Subsample 2 included 700 respondents.



Table 6

*Parallel Analysis Based on Minimum Rank Factor Analysis (PA-MRFA) of Polychoric Correlations*

Variable	Observed data % of variance	Mean of random variance	95 <sup>th</sup> percentile of random % of variance
1	51.0*	20.3	23.9
2	12.8	17.6	20.2
3	9.9	15.4	17.5
4	8.0	13.2	14.9
5	6.3	11.0	12.7
6	4.6	8.8	10.6
7	4.0	6.7	8.8
8	3.1	4.6	6.8
9	0.1	2.4	4.6
10	0.0	0.0	0.0

*Note.* Only the first 10 factors are shown. The percentage of variance relates to common variance. \* = factor retained when the observed data percentage of common variance was greater than the 95<sup>th</sup> percentile of the PA-MRFA's random datasets.

*N* = 349

Table 7

*Exploratory Factor Analysis Structure (Pattern) Coefficients for Subsample 1 (N = 349)*

Items	One-Factor Solution		Two-Factor Solution			Three-Factor Solution			
	F1	$h^2$	F1	F2	$h^2$	F1	F2	F3	$h^2$
1 A	.74	.82	.41 (-.06)	.89 ( <b>.92</b> )	.82	.43 (-.06)	.90 ( <b>.92</b> )	.26 (.04)	.82
2 A	.68	.70	.38 (-.05)	.81 ( <b>.84</b> )	.70	.41 (-.00)	.81 ( <b>.82</b> )	.19 (-.02)	.69
3 S	.74	.62	.61 (.36)	.67 ( <b>.49</b> )	.63	.63 (.37)	.66 ( <b>.46</b> )	.36 (.05)	.63
4 S	.75	.67	.68 ( <b>.49</b> )	.62 (.37)	.67	.71 ( <b>.55</b> )	.60 (.32)	.32 (-.01)	.68
5 A	.76	.72	.78 ( <b>.69</b> )	.53 (.18)	.72	.82 ( <b>.77</b> )	.51 (.12)	.36 (-.02)	.75
6 S	.65	.64	.71 ( <b>.69</b> )	.40 (.05)	.65	.71 ( <b>.71</b> )	.39 (.04)	.29 (-.03)	.57
7 A	.58	.51	.67 ( <b>.70</b> )	.31 (-.05)	.51	.68 ( <b>.69</b> )	.29 (-.08)	.34 (.05)	.50
8 EL	.48	.52	.50 ( <b>.44</b> )	.34 (.11)	.51	.40 (-.11)	.32 (.11)	.99 ( <b>1.0</b> )	.99
9 EL	.30	.31	.39 ( <b>.44</b> )	.13 (-.10)	.32	.32 (.23)	.13 (-.06)	.35 (.26)	.21
10 EL	.44	.31	.36 (.20)	.41 (.31)	.31	.35 (.14)	.41 (.32)	.24 (.09)	.31
Initial Eigenvalues	4.33		4.33	1.21		4.33	1.21	.97	
% of Common Variance	.68		.68	.14		.65	.14	.12	
Cumulative % of Common Variance	--		.68	.82		.65	.79	.91	

*Note:* A = Attachment item; S = Support item; EL = Experiential Learning item;  $h^2$  = communality. EFA structure coefficients are followed by pattern coefficients in parentheses. Pattern coefficients > .40 are in bold.

Table 8

*Fit Statistics for the Nested Confirmatory Factor Analysis Models*

Model	$\chi^2_{\text{adj.}}$	<i>df</i>	$\chi^2_{\text{DIFF}}$	$\Delta df$	<i>p</i> -value	CFI <sub>adj.</sub>	RMSEA <sub>adj.</sub>	SRMR <sub>adj.</sub>
3-factor	202.81	32	-----	-----	-----	.97	.087	.05
2-factor	369.56	34	106.22	2	<.0001	.94	.119	.07
1-factor	406.33	35	133.58	3	<.0001	.93	.123	.08

*Note.* *df* = degrees of freedom;  $\chi^2_{\text{adj.}}$  = RDWLS adjusted chi-square;  $\chi^2_{\text{DIFF}}$  = RDWLS adjusted scaled chi-square for difference tests;  $\Delta df$  = difference test degrees of freedom; CFI<sub>adj.</sub> = RDWLS adjusted comparative fit index; RMSEA<sub>adj.</sub> = RDWLS adjusted root mean square error of approximation; SRMR<sub>adj.</sub> = RDWLS adjusted standard root mean square residual. The  $\chi^2$  difference tests were between each model and the hypothesized 3-factor model. When conducting chi-square difference tests in Mplus software using RDWLS estimation, a scaling correction factor must be applied. The scaling correction factor was applied using the “DIFFTEST” command in Mplus (Muthén & Muthén, 2014).

*N* = 700

Table 9

*Unstandardized (Standardized) Factor Pattern Coefficients and Standardized Error Variances for the Three-Factor Model*

Items	Attachment	Support	Experiential Learning	Error Variances
1. [Institution] was the perfect school for people like me	1.000 (.80)	--	--	.37
2. I can't imagine a world without [institution]	.94 (.75)	--	--	.45
3. My professors at [institution] cared about me as a person	--	1.00 (.90)	--	.19
4. I had at least one professor at [institution] who made me excited about learning	--	.85 (.76)	--	.42
5. [Institution] prepared me well for life outside of college	1.02 (.82)	--	--	.33
6. While attending [institution], I had a mentor who encouraged me to pursue my goals and dreams	--	.81 (.72)	--	.48
7. [Institution] is passionate about the long-term success of its students	1.09 (.87)	--	--	.24
8. While attending [institution], I had an internship or job that allowed me to apply what I was learning in the classroom	--	--	1.00 (.54)	.71
9. While attending [institution], I worked on a project that took a semester or more to complete	--	--	.90 (.48)	.77
10. I was extremely active in extracurricular activities and organizations while attending [institution]	--	--	.80 (.43)	.82
<b>Factor Correlations</b>				
Attachment	1.0			
Support	.77	1.0		
Experiential Learning	.60	.67	1.0	
Internal Consistency Reliability ( $\omega$ )	.97	.95	.81	
Variance Extracted	.83	.81	.52	

*Note.* Unstandardized coefficients are presented followed by standardized coefficients in parentheses. All unstandardized coefficients are statistically significant at  $p < .05$ .

Table 10

*Polychoric Correlation Residuals Greater than  $|.15|$  for the One- and Two-factor CFA Models*

Item Pair	Item 1 Content	Item 2 Content	Correlation residual for one-factor model	Correlation residual for two-factor model
1 2	Perfect school for people like me	Can't imagine a world without institution	.173	-.171
1 4	Perfect school for people like me	Professor made me excited about learning	-.183	-.185
1 6	Perfect school for people like me	Mentor who encouraged me to pursue goals/dreams	-.189	-.191
8 9	Internship/job where applied what I was learning	Project that took semester or more	.167	--

*Note.* The polychoric correlation residual for Items 8 and 9 for the two-factor model was less than  $|.15|$ .

Table 11

*Fit Statistics for Full Structural Models*

Model	$\chi^2_{\text{adj.}}$	<i>df</i>	$\chi^2_{\text{DIFF}}$	$\Delta df$	<i>p</i> -value	CFI <sub>adj.</sub>	RMSEA <sub>adj.</sub>	SRMR <sub>adj.</sub>
Model A	234.745	46	-----	-----	-----	.97	.077	.047
Model B	234.995	48	5.87	2	.053	.97	.075	.048

*Note.* *df* = degrees of freedom;  $\chi^2_{\text{adj.}}$  = RDWLS adjusted chi-square;  $\chi^2_{\text{DIFF}}$  = RDWLS adjusted scaled chi-square for difference tests;  $\Delta df$  = difference test degrees of freedom; CFI<sub>adj.</sub> = RDWLS adjusted comparative fit index; RMSEA<sub>adj.</sub> = RDWLS adjusted root mean square error of approximation; SRMR<sub>adj.</sub> = RDWLS adjusted standard root mean square residual. When conducting chi-square difference tests in Mplus software using RDWLS estimation, a scaling correction factor must be applied. The scaling correction factor was applied using the “DIFFTEST” command in Mplus (Muthén & Muthén, 2014).  
*N* = 700.

Table 12

*Full Structural Model B Structure (Pattern) Coefficients, Error Variances, Disturbance Terms, and Latent Variable Variances*

Items (wording summarized)	Three-Factor Solution			Error Variance	Latent Variances
	Attachment	Support	Experiential Learning		
1 Institution was the perfect school	1.00 (.80)			.36	
2 I can't imagine a world without [institution]	.92 (.74)			.46	
3 My professors cared about me as a person		1.00 (.89)		.20	
4 At least one professor who made me excited about learning		.85 (.76)		.42	
5 [Institution] prepared me well for life outside of college	1.02 (.82)			.33	
6 I had a mentor who encouraged me to pursue goals/dreams		.82 (.73)		.47	
7 [Institution] is passionate about students' long-term success	1.08 (.87)			.25	
8 I had an internship or job that allowed me to apply what I was learning in the classroom			1.00 (.54)	.71	
9 I worked on a project that took a semester or more to complete			.86 (.46)	.79	
10 I was extremely active in extracurricular activities and organizations			.80 (.43)	.82	
Life Satisfaction	.54 (.43)	-.17 (-.16)		.90	
Work Satisfaction		.14 (.12)*	.51 (.27)	.87	
Attachment					.64
Support					.80
Experiential Learning					.29

*Note:* A = Attachment item; S = Support item; EL = Experiential Learning item. Unstandardized coefficients are presented followed by standardized coefficients in parentheses. All unstandardized coefficients, except that for support and work satisfaction, are statistically significant at  $p < .05$ .

Table 13

*Fit Statistics for Formative MIMIC Models*

Model	$\chi^2_{\text{adj.}}$	<i>df</i>	CFI <sub>adj.</sub>	RMSEA <sub>adj.</sub>
1-construct	24.72	9	.99	.05
2-construct	328.56	30	.79	.12
2-construct (free cov)	79.03	9	.95	.11
3-construct	868.65	41	.43	.17
3-construct (free cov)	84.40	8	.95	.12

*Note.* *df* = degrees of freedom;  $\chi^2_{\text{adj.}}$  = RDWLS adjusted chi-square; CFI<sub>adj.</sub> = RDWLS adjusted comparative fit index; RMSEA<sub>adj.</sub> = RDWLS adjusted root mean square error of approximation; free cov = models estimated with freely covarying formative indicators  
*N* = 700.



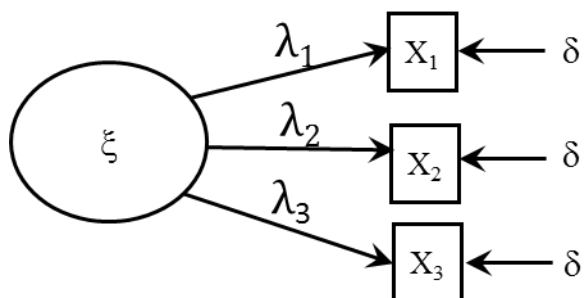
Table 14

*Unstandardized (Standardized) Factor Pattern Coefficients and Variance Explained for the One-Construct Formative Model*

Items	Attachment	$R^2$
1. [Institution] was the perfect school for people like me	<b>.175</b> (.243)	.37
2. I can't imagine a world without [institution]	-.050 (-.097)	.45
3. My professors at [institution] cared about me as a person	-.075 (-.115)	.19
4. I had at least on professor at [institution] who made me excited about learning	-.046 (-.054)	.42
5. [Institution] prepared me well for life outside of college	<b>.167</b> (.240)	.33
6. While attending [institution], I had a mentor who encouraged me to pursue my goals and dreams	<b>.071</b> (.162)	.48
7. [Institution] is passionate about the long-term success of its students	.077 (.121)	.24
8. While attending [institution], I had an internship or job that allowed me to apply what I was learning in the classroom	<b>.049</b> (.128)	.71
9. While attending [institution], I worked on a project that took a semester or more to complete	.004 (.010)	.77
10. I was extremely active in extracurricular activities and organizations while attending [institution]	.034 (.074)	.82
Life Satisfaction	<b>.932</b> (.559)	.312
Work Satisfaction	1.00 (.600)	.360
Engagement Construct		.288

*Note.* Unstandardized coefficients are presented followed by standardized coefficients in parentheses. Statistically significant unstandardized coefficients are in bold ( $p < .05$ ). All  $R^2$  values are statistically significant at  $p < .0001$ .

a) Reflective Model



b) Formative Model

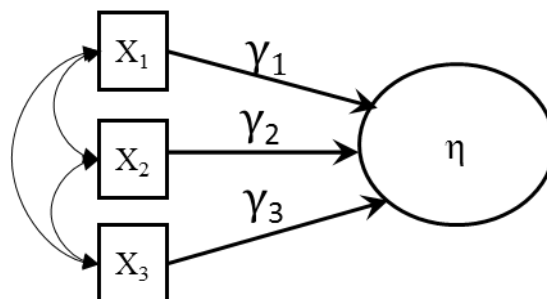
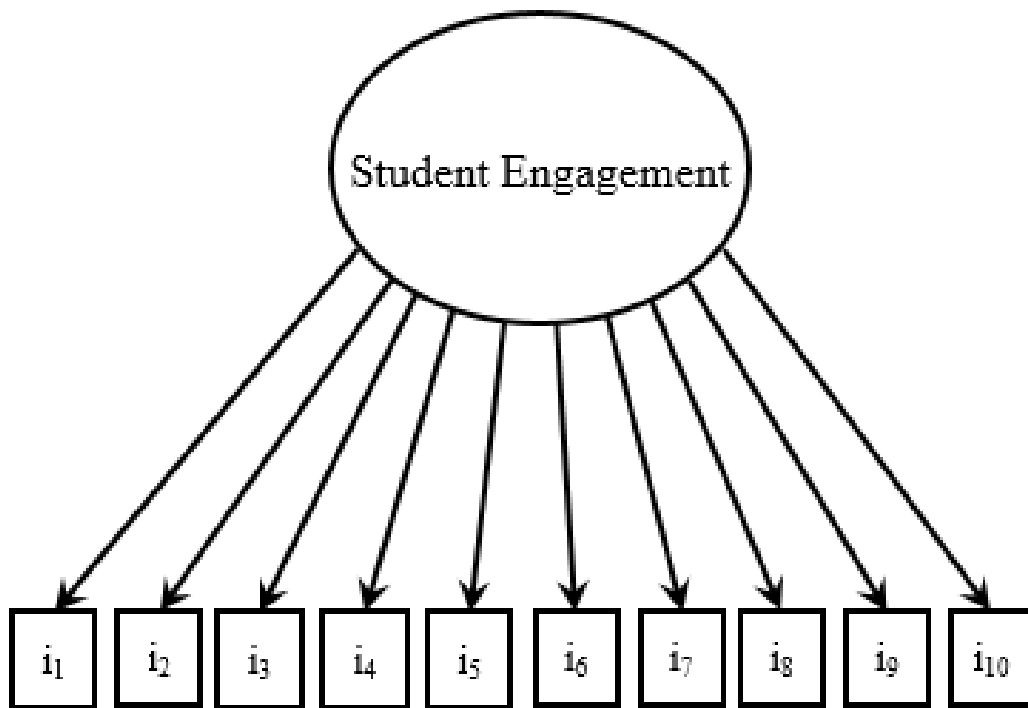
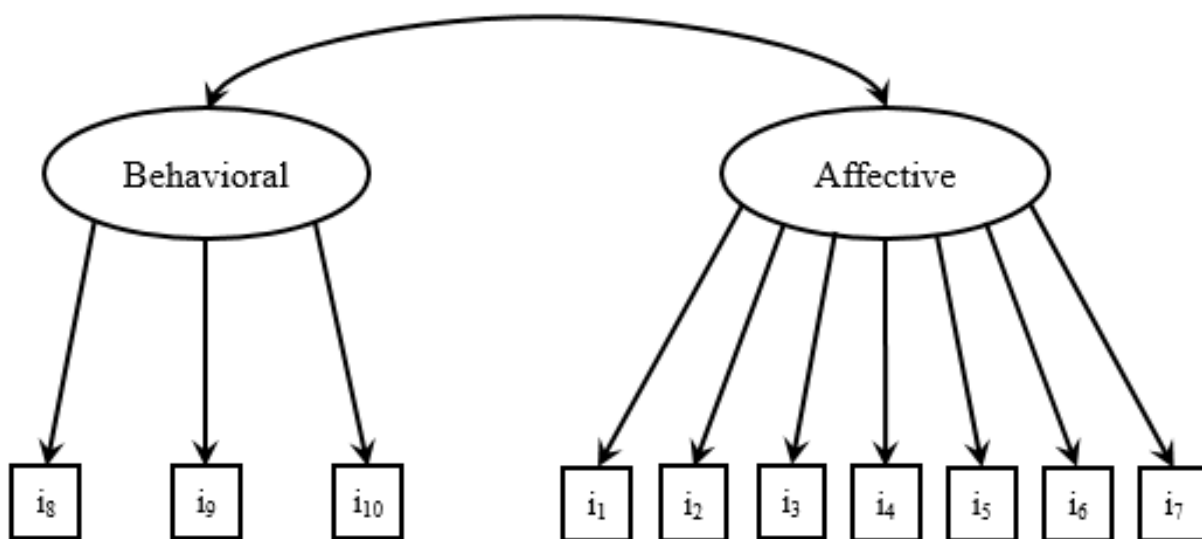


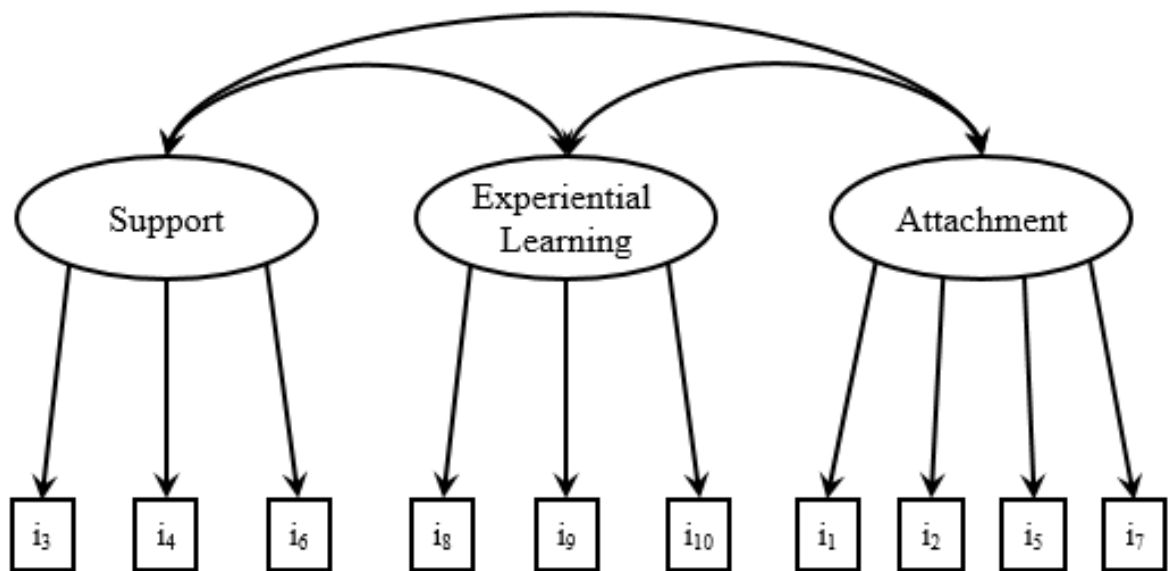
Figure 1. Example of general reflective and formative models.



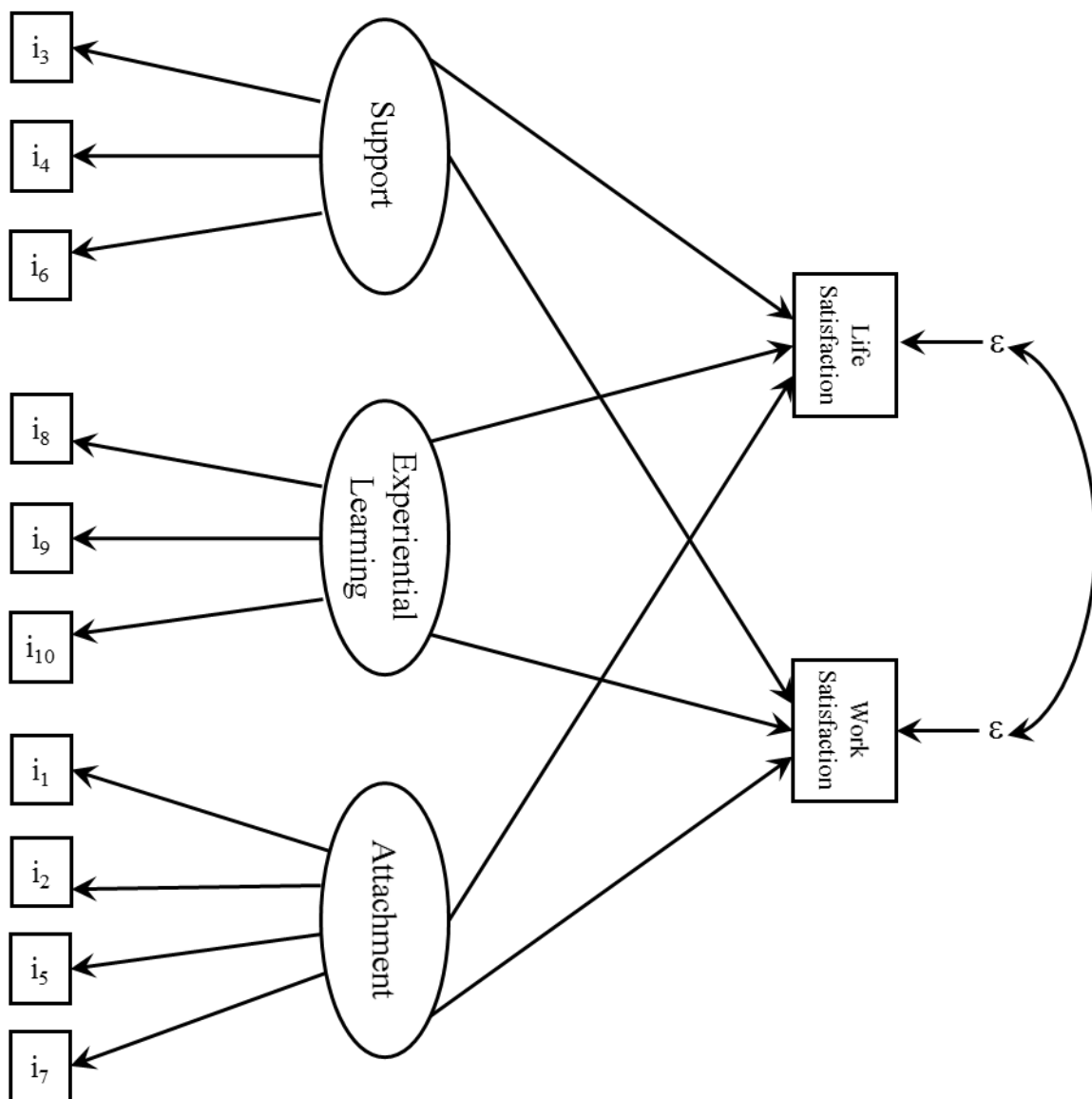
*Figure 2.* One-factor confirmatory factor analysis model.



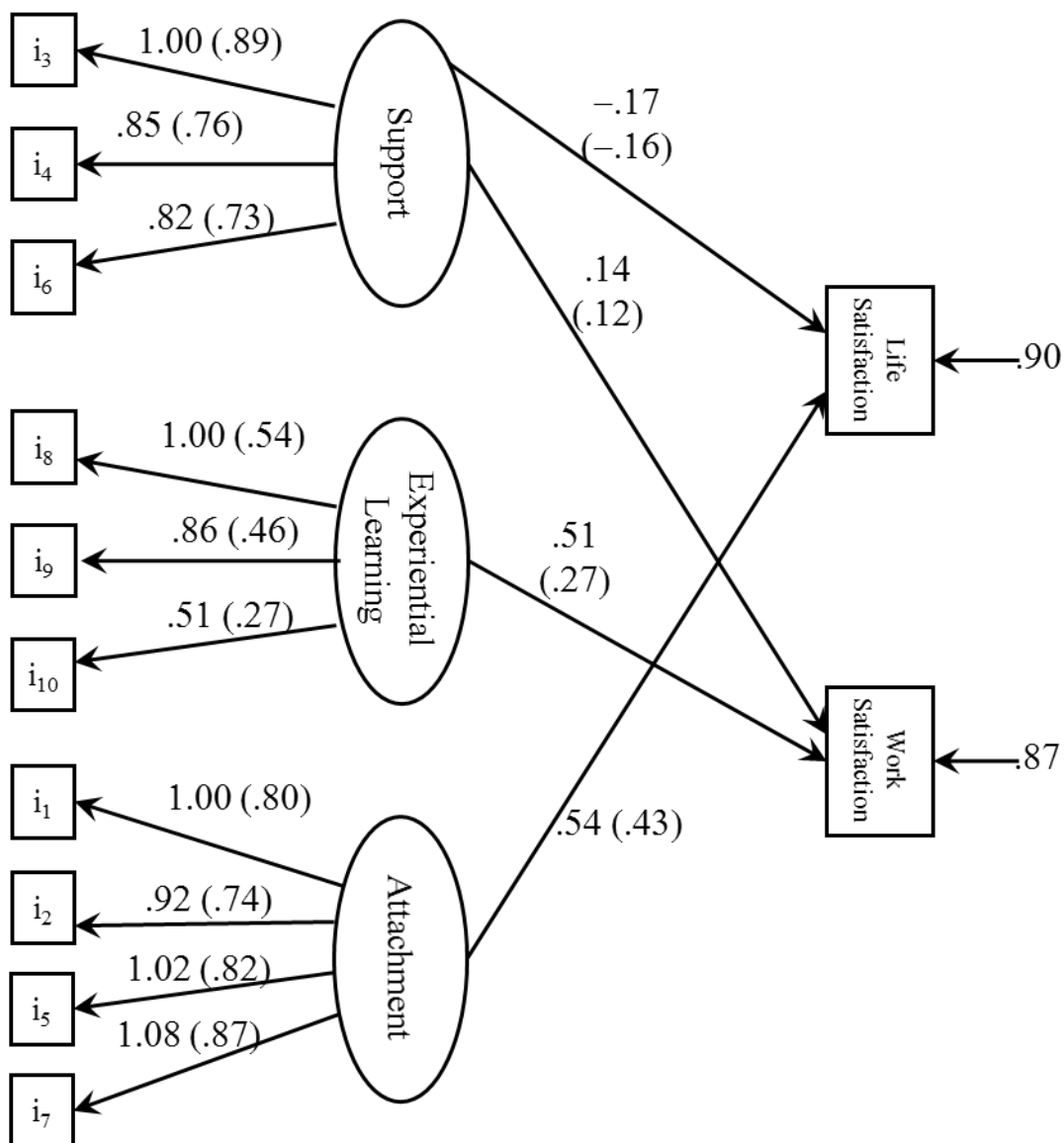
*Figure 3.* Two-factor confirmatory factor analysis model. Latent variables were allowed to freely covary.



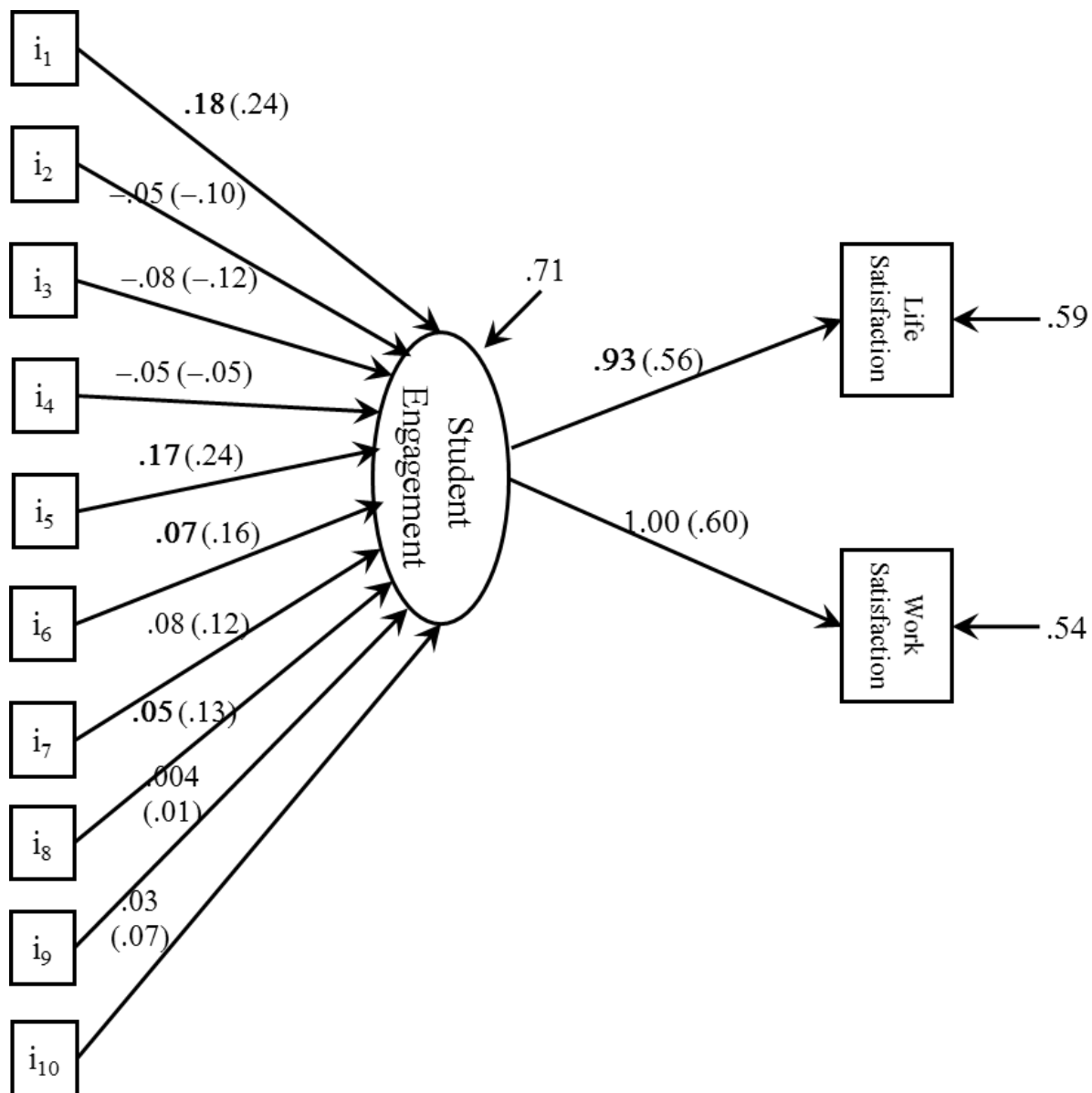
*Figure 4.* Three-factor confirmatory factor analysis model. All latent variables were allowed to freely covary.



*Figure 5.* Full Structural Model A. All paths between latent variables and outcome variables were estimated. Latent variables were allowed to freely covary. Correlations between latent variables are not depicted in the figure.



*Figure 6.* Full Structural Model B. Latent variables were allowed to freely covary. Correlations between latent variables and error variances for indicators are not depicted in the figure. Unstandardized coefficients are presented followed by standardized coefficients in parentheses. All unstandardized coefficients, except that for support and work satisfaction, are statistically significant at  $p < .05$ .



*Figure 7.* One-latent construct formative MIMIC model. All formative indicators were allowed to freely covary. Correlations between formative indicators are not depicted in the figure. Unstandardized coefficients are presented followed by standardized coefficients in parentheses. Statistically significant unstandardized coefficients are in bold ( $p < .05$ ).



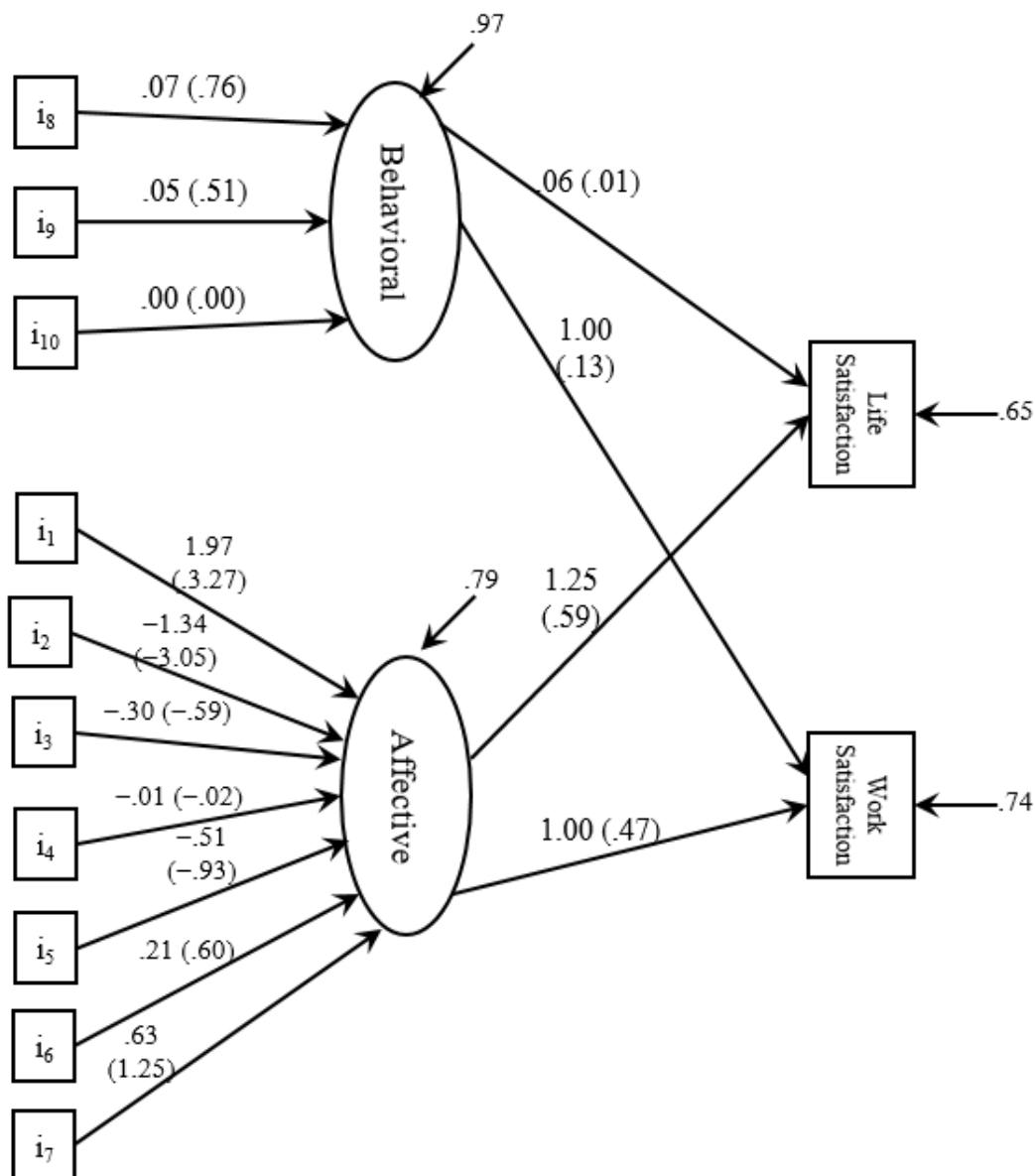
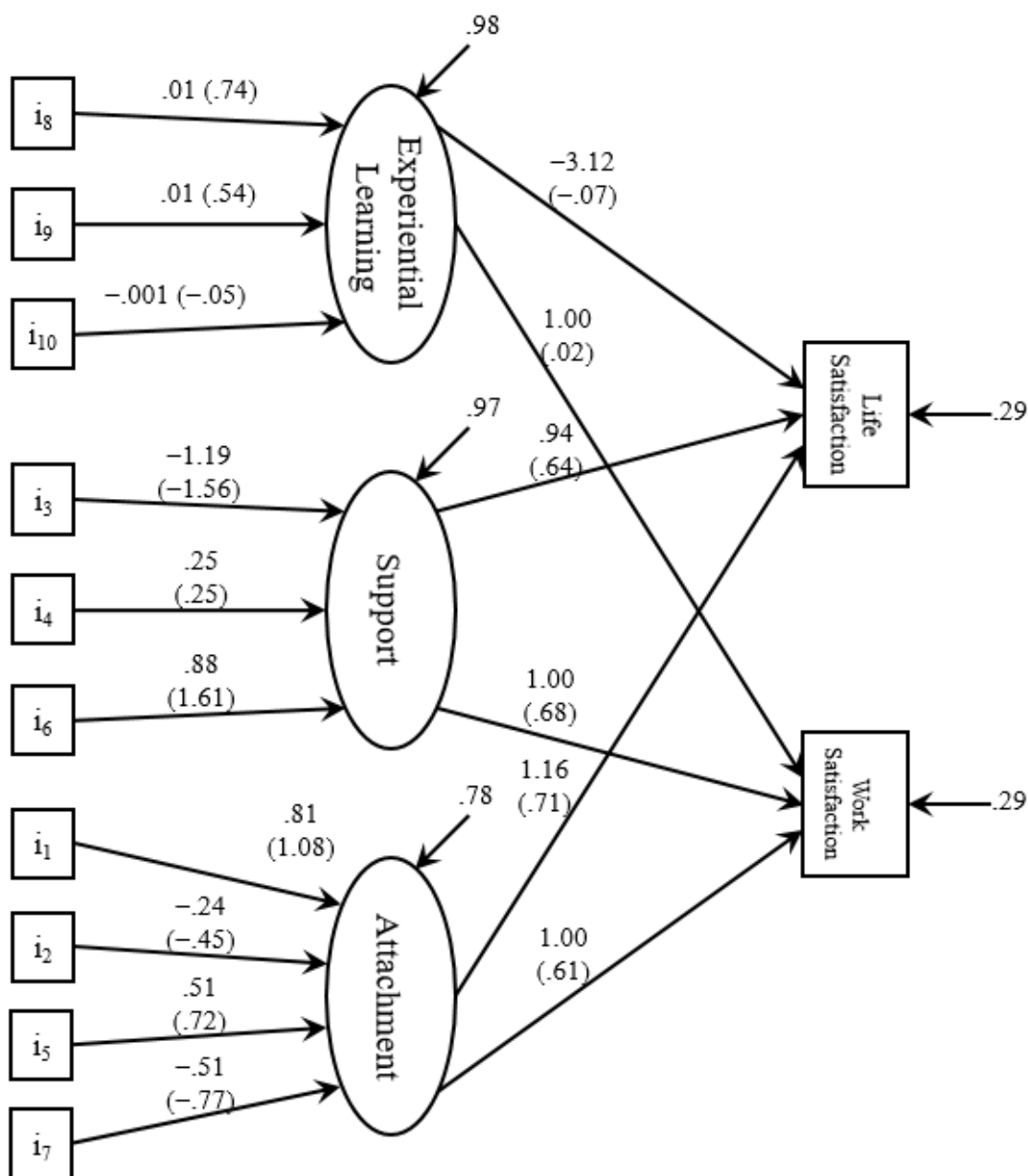


Figure 8. Two-latent construct formative MIMIC model. All formative indicators were allowed to freely covary. Correlations between formative indicators are not depicted in the figure. Unstandardized coefficients are presented followed by standardized coefficients in parentheses. Unstandardized coefficients that were not statistically significant at  $p < .05$  include those for formative indicators 4, 8, 9, and the path from the behavioral formative latent variable to the life satisfaction outcome variable.



*Figure 9.* Three-latent construct formative MIMIC model. All formative indicators were allowed to freely covary. Correlations between formative indicators are not depicted in the figure. Unstandardized coefficients are presented followed by standardized coefficients in parentheses. Unstandardized coefficients that were not statistically significant at  $p < .05$  include those for formative indicators 8, 9, 10 and the path from the experiential learning formative latent variable to the life satisfaction outcome variable.

## References

- Ahlfeldt, S., Mehta, S., & Sellnow, T. (2005). Measurement and analysis of student engagement in university classes where varying levels of PBL methods of instruction are in use. *Higher Education Research and Development, 24*(1), 5-20.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anaya, G. (1996). College experiences and student learning: The influence of active learning, college environments, and cocurricular activities. *Journal of College Student Development, 37*, 611-622.
- Anderson, J.C., & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423.
- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.
- Astin, A.W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*, 297-308.
- Astin, A.W. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.
- Association of American Colleges and Universities [AAC&U]. (2007). *College learning for the new global century: A report from the National Leadership Council for Liberal Education and America's Promise*. Washington, DC: AAC&U.

- Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research, & Evaluation, 19*(5), 1-15.
- Bagozzi, R.P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*(2), 229-237.
- Barkley, E.F. (2010). *Student engagement techniques: A handbook for college faculty*. San Francisco: Jossey-Bass.
- Baron, P., & Corbin, L. (2012). Student engagement: Rhetoric and reality. *Higher Education Research & Development, 31*(6), 759-772.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17*(1), 10-22.
- Bentler, P.M. (1998, March). Kurtosis, residuals, fit indices. Message posted to SEMNET electronic mailing list, archived at <http://bama.au.edu/archives/semnet.html>.
- Bentler, P.M., & Bonnett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.
- Bhadury, J., Dodge, K., Freedman, G., Lester, T., Merrell, G., & Ratliff, T.H. (2010). Assessment models for engagement of higher education institutions in economic development. Report commissioned by University Economic Development Association.
- Blalock, H.M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill, NC: University of North Carolina.

- Blalock, H.M. (1968). The measurement problem: A gap between the languages of theory and research. In H.M. Blalock & A.B. Blalock (Eds.), *Methodology in social research* (pp. 5-27). New York: McGraw-Hill.
- Bringle, R.G., Games, R., & Malloy, E.A. (Eds.). (1999). *Colleges and universities as citizens*. Needham Heights, MA: Allyn and Bacon.
- Bollen, K.A. (1984). Multiple indicators: Internal consistency or no necessary relationship? *Quality and Quantity*, 18, 377-385.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K.A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, 12(2), 219-228.
- Bollen, K.A. (2011). Evaluating effect, composite, and causal indicators in structural equation models. *MIS Quarterly*, 35(2), 359-372.
- Bollen, K.A., & Davis, W.R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 498-522. (Reprinted from Bollen, K.A., & Davis, W. (1994). *Causal indicator models: Identification, estimation, and testing*. Paper presented at the Annual Meeting of the Association for the American Sociological Association, Miami, FL.)
- Bollen, K.A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), 305-314.
- Borsboom, D., Mellenbergh, G.H., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203-219.

- Bowen, S. (2005). Engaged learning: Are we all on the same page? *Peer Review*, 4-7.
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.). *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.
- Burt, R.S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods and Research*, 5, 3-52.
- Butler, J.M. (2011). Using standardized tests to assess institution-wide student engagement. In R.L. Miller, E. Amsel, B. Kowalewski, B.M. Beins, K.D. Keith, & B.F. Peden (Eds.). *Promoting student engagement, volume 1: Programs, techniques and opportunities*. Syracuse, NY: Society for the Teaching of Psychology. Retrieved from <http://teachpsych.org/ebooks/pse2011/vol1/index.php>
- Campbell, C.M., & Cabrera, A.F. (2011). How sound is NSSE?: Investigating the psychometric properties of NSSE at a public, research-extensive institution. *The Review of Higher Education*, 35(1), 77-103.
- Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.
- Carini, R.M., Kuh, G.D., Klein, S.P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education*, 47, 1-32.
- Cattell, R.B. (1966). Scree test for number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Chapman, E. (2003). Alternative approaches to assessing student engagement rates. *Practical Assessment, Research & Evaluation*, 8(13). Retrieved from <http://pareonline.net/getvn.asp?v=8&n=13>

- Chickering, A.W., & Gamson, Z.F. (1987, March). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, 3-7.
- Chin, W.W., Peterson, R.A., Brown, S.P. (2008). Structural equation modeling in marketing: Some practical reminders. *Journal of Marketing Theory and Practice*, 16(4), 287-298.
- Chou, C., & Bentler, P.M. (1995). Estimates and test in structural equation modeling. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (p. 37-55).
- Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education*, 11, 25-36.
- Cole, J., & Dong, Y. (2013). *Confirmatory factor analysis of the BCSSE scales*. Bloomington, IN: Center for Postsecondary Research. Retrieved from <http://bcsse.indiana.edu/pdf/BCSSE%20Psychometric%20Report.pdf>
- College Student Experiences Questionnaire [CSEQ]. (2007). *General Information*. Retrieved from [http://cseq.indiana.edu/cseq\\_generalinfo.cfm](http://cseq.indiana.edu/cseq_generalinfo.cfm)
- Coltman, T., Devinney, T.M., Midgley, D.F., & Venaik, S. (2008). Formative versus reflective measurement models: two applications of formative measurement. *Journal of Business Research*, 61(12), 1250-1262.
- Cruce, T.M., Wolniak, G.C., Seifert, T.A., & Pascarella, E.T. (2006). Impacts of good practices on cognitive development, learning orientations, and graduate degree plans during the first year of college. *Journal of College Student Development*, 47(4), 365-383.

- Diamantopoulos, A. (2006). The error term in formative measurement models: Interpretation and modeling implications. *Journal of Modeling in Management, 1*, 7-17.
- Diamantopoulos, A., Riefler, P., & Roth, K.P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203-1218.
- Diamantopoulos, A., & Sigauw, J.A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management, 17*, 263-282.
- Diamantopoulos, A., & Winklhofer, H.M. (2001). Index constriction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269-277.
- Diener, E., Kahneman, D., Tov, W., & Arora, R. (2009). Income's differential influence on judgments of life versus affective wellbeing. *Assessing Wellbeing*. Oxford, UK: Springer.
- Donald, J.G. (1997). *Improving the environment for learning: Academic leaders talk about what works*. San Francisco, CA: Jossey-Bass.
- Edgerton, R. (1997). *Education white paper* (unpublished manuscript). Retrieved from [http://www.faculty.umb.edu/john\\_saltmarsh/resources/Edgerton%20Higher%20Education%20White%20Paper.rtf](http://www.faculty.umb.edu/john_saltmarsh/resources/Edgerton%20Higher%20Education%20White%20Paper.rtf)
- Edwards, J.R. (2011). The fallacy of formative measurement. *Organizational Research Methods, 14*(2), 370-388.
- Edwards J.R., & Bagozzi, R.P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*, 155-174.



- Ewell, P.T. (2010). The US National Survey of Student Engagement (NSSE). In D. Dill & M. Beerkens (Eds.), *Public policy for academic quality: Analyses of innovative policy instruments*. New York: Springer.
- Finney, S.J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation models. In G.R. Hancock & R.O. Mueller (Eds.) *A second course in structural equation modeling*. Greenwich, CT: Information Age Publishing, Inc.
- Gabriel, K.F. (2008). *Teaching underprepared students: Strategies for promoting success and retention in higher education*. Sterling, VA: Stylus.
- Gallup. (2014a). *How does the U.S. Gallup-Healthways Well-Being Index work?* Retrieved from <http://www.gallup.com/175196/gallup-healthways-index-methodology.aspx>
- Gallup. (2014b). *How does the Gallup-Purdue Index work?* Retrieved from [http://www.gallup.com/174167/temp-methodology-gallup-purdue-index-methodology.aspx?g\\_source=&g\\_medium=&g\\_campaign=tiles](http://www.gallup.com/174167/temp-methodology-gallup-purdue-index-methodology.aspx?g_source=&g_medium=&g_campaign=tiles)
- Gallup. (2016). *Gallup employee engagement: The engaged workplace*. Retrieved from <http://www.gallup.com/services/190118/engaged-workplace.aspx>
- Gallup-Purdue. (2014). *Great jobs, great lives: The 2014 Gallup-Purdue Index report*. Retrieved from <http://products.gallup.com/168857/gallup-purdue-index-inaugural-national-report.aspx>
- Gallup-Purdue. (2015). *Great jobs, great lives. The relationship between student debt, experiences, and perceptions of college worth: Gallup-Purdue Index 2015 report*. Retrieved from <http://www.gallup.com/services/185888/gallup-purdue-index-report-2015.aspx>

- Gilchrist, D., & Oakleaf, M. (2012). *An essential partner: The librarian's role in student learning assessment* (NILOA Occasional Paper No. 14). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Gonyea, R.M., Kish, K.A., Kuh, G.D., Muthiah, R.N., & Thomas, A.D. (2003). *College Student Experiences Questionnaire: Norms for the Fourth Edition*. Bloomington, IN: Indiana University Center for Postsecondary Research, Policy, and Planning.
- Gordon, J., Ludlum, J., & Hoey, J.J. (2008). Validating NSSE against student outcomes: Are they related? *Research in Higher education*, 49, 19-39.
- Grace, J.B., & Bollen, K.A. (2008). Representing general theoretical concepts in structural equation models: The role of composite variables. *Environmental and Ecological Statistics*, 15(2), 191-213.
- Grunwald, H., & Peterson, M.W. (2003). Factors that promote faculty involvement in and satisfaction with institutional and classroom student assessment. *Higher Education*, 44(2), 173-204.
- Handelsman, M.M., Briggs, W.L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *Journal of Educational Research*, 98, 184-191.
- Harman, H.H. (1976). *Modern factor analysis* (3<sup>rd</sup> ed.). Chicago, IL: University of Chicago Press.
- Harper, S.R., Carini, R., Bridges, B., & Hayek, J. (2004). Gender differences in student engagement among African American undergraduates at historically Black colleges and universities. *Journal of College Student Development*, 45, 271-284.

- Harper, S.R., & Quaye, S.J. (2009). Beyond sameness with engagement and outcomes for all: An introduction. In S. Harper & S. Quaye (Eds.), *Student engagement in higher education* (pp. 1-15). New York: Routledge.
- Harper, S.R., & Quaye, S.J. (2015). Making engagement equitable for students in U.S. higher education. In S.J. Quaye & S.R. Harper (Eds.), *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations* (Second edition). New York: Routledge.
- Hathcoat, J.D., & Meixner, C. (2015). Pragmatism, factor analysis, and the conditional incompatibility thesis in mixed methods research. *Journal of Mixed Methods Research*, 1-17.
- Heise, D.R. (1972). Employing nominal variables, induced variables, and block variables in path analysis. *Sociological Methods & Research*, 1, 147-173.
- Hood, S.B. (2013). Psychological measurement and methodological realism. *Erkenntnis*, 78(4), 739-761.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Howell, R.D., Breivik, E., & Wilcox, J.B. (2007). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205-218.
- Hu, L. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.

- Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
- Hu, S., & Kuh, G.D. (2003). Maximizing what students get out of college: Testing a learning productivity model. *Journal of College Student Development, 44*, 185-203.
- Hu, S., Scheuch, K., Schwartz, R., Gayles, J., & Li, S. (2008). *Revisiting undergraduate education: Engaging college students in research and creative activities*. ASHE Higher Education Report, Vol. 33, no. 4. San Francisco: Jossey-Bass.
- Hu, S., & Wolniak, G.C. (2010). Initial evidence on the influence of college student engagement on early career earnings. *Research in Higher Education, 51*, 750-766.
- Hu, S., & Wolniak, G.C. (2013). College student engagement and early career earnings: Differences by gender, race/ethnicity, and academic preparation. *The Review of Higher Education, 36*(2), 211-233.
- IBM Corporation. (2014). IBM SPSS Statistics for Windows (Version 23.0) [Computer Software]. Armonk, NY: IBM Corp.
- Jarvis, C.B., MacKenzie, S.B., & Podsakoff, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*, 199-218.
- Joliffe, I.T. (2002). *Principal components analysis* (2nd ed.) New York: Springer-Verlag.
- Jöreskog, K.G., & Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*(351), 631-639.

- Kane, M.T. (1992). An argument-based approach to validation. *Psychological Bulletin*, *112*, 527-535.
- Kane, M.T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319-342.
- Kline, R.B. (2011). *Principles and practice of structural equation modeling* (3rd ed). New York: Guilford Press.
- Kuh, G.D. (2008a). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington, DC: American Association of Colleges and Universities.
- Kuh, G.D. (2008b). Why integration and engagement are essential to effective educational practice in the twenty-first century. *Peer Review*, *10*(4), 27-28.
- Kuh, G.D. (2009). The National Survey of Student Engagement: Conceptual and empirical foundations. In R.M. Gonyea & G.D. Kuh (Eds.), *Using NSSE in institutional research. New Directions for Institutional Research*, *141*, 5-20.
- Kuh, G.D., Cruce, T.M., Shoup, R., Kinzie, J., & Gonyea, R.M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education*, *79*(5), 540-563.
- Kuh, G.D., Hayek, J.C., Carini, R.M., Ouimet, J.A., Gonyea, R.M., & Kennedy, J. (2001). *NSSE technical and norms report*. Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.
- Kuh, G.D., Nelson Laird, T.F., & Umbach, P.D. (2004). Aligning faculty activities and student behavior: Realizing the promise of greater expectations. *Liberal Education*, *90*(4), 24-31.

- LaNasa, S.M., Cabrera, A.F., & Trangsrud, H. (2009). The construct validity of student engagement: A confirmatory factor analysis approach. *Research in Higher Education, 50*, 315-332.
- LaNasa, S.M., Olson, E., & Alleman, N. (2007). The impact of on-campus student growth on first-year student engagement and success. *Research in Higher Education, 48*(8), 941-966.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635-694.
- Lopez, S.J., Agrwal, S., & Calderon, V.J. (2010). *The Gallup Student Poll Technical Report*. Gallup, Inc. Retrieved from <http://www.gallup.com/services/177095/gallup-student-poll-technical-report.aspx>
- Lorenzo-Seva, U., & Ferrando, P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88-91.
- MacCallum, R., & Browne, M. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114*(3), 533-541.
- Mandernach, B.J. (2015). Assessment of student engagement in higher education: A synthesis of literature and assessment tools. *International Journal of Learning, Teaching, and Educational Research, 12*(2), 1-14.
- Marsh, H. W., Hau, K-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers to overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.

- Marti, C.N. (2009). Dimensions of student engagement in American community colleges: Using the community college student report in research and practice. *Community College Journal of Research and Practice*, 33, 1-24.
- McCormick, A.C., Kinzie, J., & Gonyea, R.M. (2013). Student engagement: Bridging research and practice to improve the quality of undergraduate education. In M.B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 28, pp. 47-92). New York: Springer.
- McCormick, A.C., & McClenney, K. (2012). Will these trees *ever* bear fruit? A response to the special issue on student engagement. *The Review of Higher Education*, 35(2), 307-333.
- Merwin, J.C. (1969). Historical review of changing concepts of evaluation. In R.L. Tyler (Ed.), *Educational evaluation: New roles, new methods*. The sixty-eighth yearbook of the National Society for the Study of Education, Part II. Chicago, IL: University of Chicago Press.
- Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, 89, 575-588.
- Muthén, L. K., & Muthén, B. O. (1998-2014). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- National Commission on the Future of Higher Education. (2006). *A test of leadership: Chartering the future of U.S. higher education*. Washington, DC: US Department of Education.
- National Survey of Student Engagement [NSSE]. (2001). *Our origins and potential*. Retrieved from <http://nsse.indiana.edu/html/origins.cfm>

- National Survey of Student Engagement [NSSE]. (2005). *Student engagement: Exploring different dimensions of student engagement*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- National Survey of Student Engagement [NSSE]. (2009). *NSSE timeline: 1998-2009*. Retrieved from [http://nsse.indiana.edu/pdf/NSSE\\_Timeline.pdf](http://nsse.indiana.edu/pdf/NSSE_Timeline.pdf)
- National Survey of Student Engagement [NSSE]. (2016). *About NSSE*. Retrieved from <http://nsse.indiana.edu/html/about.cfm>
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4), 443-460.
- Pace, C.R. (1980). Measuring the quality of student effort. *Current Issues in Higher Education*, 2, 10-16.
- Pace, C.R. (1982). *Achievement and the quality of student effort*. Washington, DC: National Commission on Excellence in Education.
- Pace, C.R. (1990). *The undergraduates: A report of their activities and college experiences in the 1980s*. Los Angeles, CA: Center for the Study of Evaluation, UCLA Graduate School of Education.
- Pascarella, E.T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 1). New York: Agathon.
- Pascarella, E.T., Seifert, T.A., & Blaich, C. (2008). *Validation of the NSSE benchmarks and deep approaches to learning against liberal arts outcomes*. Paper presented at



the Annual Meeting of the Association for the Study of Higher Education,  
Jacksonville, FL.

- Pascarella, E.T., Seifert, T.A., & Blaich, C. (2010). How effective are the NSSE benchmarks in predicting important educational outcomes? *Change*, 42(1), 16-22.
- Pascarella, E.T., & Terenzini, P.T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.
- Pascarella, E.T., & Terenzini, P.T. (2005). *How college affects students: A third decade of research* (Vol. 2). San Francisco: Jossey-Bass.
- Paulsen, M.B. (Ed.). (2013). *Higher education: handbook of theory and research* (Vol. 28). New York: Springer.
- Porter, S.R. (2011). Do college student surveys have any validity? *The Review of Higher Education*, 35, 45-76.
- Posner, B.Z. (2004). A leadership instrument for students: Updated. *Journal of College Student Development*, 45, 443-456.
- Quaye, S.J. & S.R. Harper (2015). *Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations* (Second edition). New York: Routledge.
- Rath, T., & Harter, J. (2011). The economics of well-being. *Gallup Consulting*. Retrieved from [http://www.gallup.com/file/services/177050/The\\_Economics\\_of\\_Wellbeing.pdf](http://www.gallup.com/file/services/177050/The_Economics_of_Wellbeing.pdf)
- Rath, T., & Harter, J. (2010). *Well-being: The five essential elements*. New York: Gallup Press.

- Ray, J., & Kafka, S. (2014). Life in college matters for life after college. *Gallup*.  
Retrieved from <http://www.gallup.com/poll/168848/life-college-matters-life-college.aspx>
- Rocconi, L.M. (2011). The impact of learning communities on first year students' growth and development in college. *Research in Higher Education*, 52, 178-193.
- Rodgers, J.L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, 65(1), 1-12.
- Saltmarsh, J., & Hartley, M. (Eds.). (2011). *"To serve a larger purpose": Engagement for democracy and the transformation of higher education*. Philadelphia, PA: Temple University Press.
- Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent variable analysis*. Thousand Oaks, CA: Sage.
- Schmaling, K.B., & Guy, S.M. (2014). Associations among National Survey of Student Engagement benchmarks, academic achievement, and life satisfaction as alumni. *International HETL Review*, 4(10). Retrieved from <https://www.hetl.org/academic-articles/associations-among-national-survey-of-student-engagement-benchmarks-academic-achievement-and-life-satisfaction-as-alumni>
- Sears, L.E., Agrawal, S., Sidney, J.A., Castle, P.H., Rula, E.Y., Coberley, C.R., ...  
Harter, J.K. (2014). The well-being 5: Development and validation of a diagnostic instrument to improve population well-being. *Population Health Management*, 17(6), 357-365.

- Skinner, E.A., & Belmont, J. (1993). Motivation in the classroom: Reciprocal effects of teach behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571-581.
- Smith, K.A., Sheppard, S.D., Johnson, D.W., & Johnson, R.T. (2005). Pedagogies of engagement: Classroom-based practices. *Journal of Engineering Education, 94(1)*, 87-102.
- Steenkamp, J.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.
- Strange, C.C., & Banning, J.H. (2001). *Education by design: Creating campus learning environments that work*. San Francisco, CA: Jossey-Bass, Inc.
- Swerdzewski, P., Miller, B.J., & Mitchell, R. (2007). *Investigating the validity of the National Survey of Student Engagement*. Paper presented at the Northeastern Educational Research Association, Rocky Hill, CT.
- Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson Education, Inc.
- Timmerman, M.E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16(2)*, 209-220.
- Tinto, V. (1975). Dropouts from higher education: A theoretical synthesis of the recent literature. *Review of Educational Research, 45*, 89-125.
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed). Chicago: University of Chicago Press.
- Trowler, V. (2010). *Student engagement literature review*. York, United Kingdom: The Higher Education Academy.

- Trowler, V., & Trowler, P. (2010). *Student engagement evidence summary*. York, United Kingdom: The Higher Education Academy.
- Wang, X., French, B.F., & Clay, P.F. (2015). Convergent and discriminant validity with formative measurement: A mediator perspective. *Journal of Modern Applied Statistical Methods*, *14*(1), 83-106.
- Williams, J.M. (2007). *College Student Experiences Questionnaire assessment program*. Presentation at the Annual Meeting of the Association for Institutional Research, Kansas City, MO.
- Yu, C.-Y. & Muthén, B. (2002). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Technical report.
- Zhao, C., & Kuh, G.D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education*, *45*(2), 115-138.