Student Engagement and Learning Gains: Self-reports, Direct Measures, and Instrument Specificity

Christopher Kerr

Follow this and additional works at: https://commons.lib.jmu.edu/master201019
Part of the Education Commons, and the Psychology Commons

Recommended Citation
https://commons.libjmu.edu/master201019/546

This Thesis 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 Masters Theses by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.
Student Engagement and Learning Gains: 
Self-reports, Direct Measures, and Instrument Specificity

Christopher Shane Kerr II

A thesis submitted to the Graduate Faculty of
JAMES MADISON UNIVERSITY
In
Partial Fulfillment of the Requirements
for the degree of
Master of Arts

Department of Graduate Psychology

August 2018

FACULTY COMMITTEE:
Committee Chair: Dr. Keston H. Fulcher
Committee Members/ Readers:
Dr. Dena Pastor
Dr. Fletcher Linder
Table of Contents

List of Tables .................................................................................................................. v
List of Figures .................................................................................................................. vi
Abstract .......................................................................................................................... vii
I. Chapter One: Introduction ......................................................................................... 1
  Overview ....................................................................................................................... 1
II. Chapter Two: Literature Review ............................................................................. 10
  Origin of Engagement Research ................................................................................ 10
  Conceptualizing Engagement ..................................................................................... 12
  Measuring Engagement ............................................................................................... 15
  College Student Engagement Questionnaire (CSEQ) ............................................. 15
  National Survey of Student Engagement .................................................................. 16
  NSSE validity ............................................................................................................. 18
  Additional Measurement Issues .................................................................................. 28
  Student Learning ......................................................................................................... 29
  Measuring outcomes ................................................................................................. 30
  Direct/Indirect measures ............................................................................................ 31
Self-Reported Learning Gains as Proxies for Direct Measures of Learning Gains ......... 31
  Engagement and its Relationship to Learning Outcomes ......................................... 33
  GPA ............................................................................................................................ 34
  Engagement-Knowledge/Ability ............................................................................... 35
  Engagement-SRLG ...................................................................................................... 37
  Engagement-DMLG ..................................................................................................... 38
  Methodological Concerns of Engagement-Outcome Research ............................... 39
  Alignment of Measures .............................................................................................. 44
  Summary ..................................................................................................................... 46
III. Chapter Three: Methodology ............................................................................... 49
  Participants .................................................................................................................. 49
  Measures ...................................................................................................................... 50
  NSSE: Perceived Quantitative Learning Gains (General Self-reported Learning) .......... 50
  Quantitative Reasoning Self-Reported Learning Gain Scale (Specific Self-Reported Learning) .................................................................................................................. 51
  Quantitative Reasoning-9 (Direct Measure) .............................................................. 52
  Engagement Measures ............................................................................................... 53
  NSSE: Engagement (General Engagement) ............................................................... 54
  Adapted NSSE: Quantitative Reasoning Engagement (Specific Engagement) ....... 54
Student Opinion Survey ................................................................. 55
Procedures .................................................................................. 56
Missing Data .................................................................................. 56
Motivation Filtering ....................................................................... 57
Analysis .......................................................................................... 57
RQ1: How does students’ general self-reported learning differ from specific self-reported learning? .................................................. 58
RQ2: Do students’ direct measures of learning gains differ depending on their Self-reported learning gains? ............................................. 59
RQ3: To what degree do measures of general engagement and quantitative engagement differ from each other? ........................................... 59
RQ4: What are the relationships between engagement and quantitative reasoning learning gains? .......................................................... 60
RQ5: Does general engagement predict quantitative reasoning learning gains as well as specific engagement? ............................................. 60
RQ6: Do general and specific engagement predict learning differently for each measure of learning? .......................................................... 60
RQ7: Does the relationship between the most commonly used measures of learning and engagement (general SRLG and general engagement) differ from the relationship using the ideal forms of the engagement and learning measures (DMLG and specific engagement)? .................................................. 61
IV. Chapter Four: Results .................................................................. 63
Research Question One: General vs. Specific Self-reported Learning Gains ...... 63
Descriptives .................................................................................. 63
Analysis .......................................................................................... 63
Research Question Two: Direct-measured Learning Gains vs. General and Specific Self-reported Learning Gains ............................................. 64
Descriptives .................................................................................. 64
Analysis .......................................................................................... 64
Research Question Three: General Engagement vs. Specific Engagement ........ 65
Descriptives .................................................................................. 65
Analysis .......................................................................................... 65
Research Question Four: Relationships Between Engagement and Learning..... 66
Analysis .......................................................................................... 66
Research Question Five: General Engagement vs. Specific Engagement in Predicting Learning .......................................................... 66
Analysis .......................................................................................... 66
Research Question Six: Differences Between Measures of Learning in Their Relation to Engagement ................................................................. 67
Analysis ........................................................................................................ 67

Research Question Seven: Engagement-Learning Relationships Using Common vs. Ideal Methodology .......................................................... 69
Analysis ........................................................................................................ 69

V. Chapter Five: Discussion ............................................................................ 71

Research Question One: General Self-reported Learning and Specific Self-reported Learning .............................................................................. 71
Research Question Two: Direct-measured Learning Gains vs. General and Specific Self-reported Learning Gains .................................................. 72
Research Question Three: General Engagement vs. Specific Engagement ...... 73
Research Question Four: Relationships Between Engagement and Learning ..... 74
Research Question Five: General Engagement vs. Specific Engagement in Predicting Learning ................................................................................ 75
Research Question Six: Differences Between Measures of Learning in their Relation to Engagement ................................................................. 76
Research Question Seven: Engagement-Learning Relationships Using Common vs. Ideal .................................................................................. 78
Implications ..................................................................................................... 80
Limitations and Future Directions ................................................................. 82

Conclusion .................................................................................................... 82

Tables ............................................................................................................. 86
Figures .......................................................................................................... 96
References .................................................................................................... 98
List of Tables

Table 1. Themes and indicators of the NSSE 2.0 86
Table 2. Learning outcomes measures by the NSSE 86
Table 3. Original and adapted questions from the National Survey of Student Engagement 87
Table 4. Demographics 90
Table 5. Perceived Quantitative Learning Gains 90
Table 6. Natural World Self-Reported Learning Gain Scale 91
Table 7. Quantitative Reasoning 9 Example Items 91
Table 8. NSSE Indicators 92
Table 9. Example NSSE General Engagement Items 92
Table 10. Example NSSE Specific Engagement Items 92
Table 11. Student Opinion Scale: Example Items 93
Table 12. Table comparing students QR-9 gains across responses levels to NSSE SRLG 93
Table 13. Example ANOVA table comparing students QR-9 gains across responses levels to specific SRLG 94
Table 14. Correlations between engagement and measures of learning on general engagement 94
Table 15. Correlations between engagement and measures of learning on QR specific engagement 95
List of Figures

Figure 1. Diagram illustrating student learning as influenced by indirect factors 96

Figure 2. Institutional factors and student engagement influence on learning 97
Abstract

The assumption that engagement leads to student learning is fundamental in higher education. Engagement is often used by educational institutions as an indicator of student learning. However, research has found moderate to weak relationships between engagement and learning. This study explored the influence that methods used to measure learning and engagement may have on the relationships observed between the two. More specifically, this study considered differences between self-reported measures of learning and direct-measure change scores in their relationship to engagement. Additionally, this study tested the influence that specificity of engagement measures has on observed engagement-learning relationships. Quantitative reasoning was selected as the learning area in which these relationships were examined.

Three hundred and thirteen participants were randomly assigned to one of two testing conditions. One group completed a set of questions taken from the National Survey of Student Engagement. The other group completed the same set of engagement questions that were narrowed to reflect quantitative reasoning. All students completed the same measures of learning, namely a general self-reported measure of learning gains, a specific measure of self-reported learning gains, and a direct measure of learning gains calculated using the Quantitative Reasoning-9 test.

First, self-reported and direct measures of learning were compared to each other. Then, relationships between each measure of engagement and each measure of learning were calculated and compared. Results of the analyses indicated that student self-reports of learning (SRLG) were not representative of their direct measured learning gains (DMLG). Comparison of correlations revealed no difference in the relationship of
general and specific engagement to any measure of learning. Additionally, analyses revealed differences between SRLG and DMLG in their relationship to engagement. These results emphasize the importance of practicing caution when using SRLG while studying the relationships between engagement and learning. Ideally, researchers should consider only using measures of SRLG which have been found to lead to comparable results as well validated DMLG. Further implications and suggestions for future research are also provided.
CHAPTER ONE

Introduction

Overview

The past three decades have been a time of soul searching for the American higher education system. Historically, college degrees have been valued by the public, employers, and government for their financial and intrinsic benefits (Bok, 2009). However, over the past several decades attitudes about the value of higher education have shifted toward skepticism. Increasingly, policymakers, students, and employers have expressed concern that college students’ academic achievement may be falling short of expectations (Alexander, 2000; Ewell, 2002; 2008; Venezia, Callan, Kirst, & Usdan, 2005). In many cases, the blame for low student achievement has fallen on the educational institutions. As concerns about the value of higher education grew, so did policymakers’ demands for greater accountability in higher education (Alexander, 2000; Ewell, 2002; 2008; 2009; Pike, 1996). This placed pressure on higher education to demonstrate its value (Bok, 2009; Ewell, 2002; 2009). During this time, some within academe recognized the potential of assessment as a means of demonstrating and improving positive student outcomes (Ewell, 2002).

In the early 1980s, the assessment movement emerged in response to the mounting external and internal pressures to demonstrate student learning (Ewell, 2002; 2008; Finn & Zimmer, 2012). Prior to the 1980s, conducting assessment in higher education was a matter of choice, used primarily to obtain information for program improvement (Ewell, 2008; Kuh, 2001). However, assessment soon became an expectation as with increasing calls for better curriculum and student outcomes (Ewell, 2002). Around the same time, policymakers began to realize assessment’s utility in
measuring the return on taxpayer dollars (Ewell, 2009).

As assessment became an expectation it began to serve two roles: to improve student learning and as a means of accountability to officials and policymakers (Erwin, 2003; Ewell, 2002; 2009; McCormick, 2008; Pike, 1996). These same roles persist in higher education assessment today (Ewell, 2009).

For either improvement or accountability, two basic questions dominate discussion: What do students know upon graduating? And how much do students learn from their college experience (Ewell, 2002)? Answering these questions requires the ability to measure and demonstrate student competency and learning, often referred to as student learning outcomes assessment (SLOs; Ewell, 2009; Pike, 1996).

Early in the assessment movement, institutions struggled to measure student learning. Because there were few instruments designed to measure SLOs, tests such as the ACT and GRE were poorly adapted to and used to measure SLOs (Ewell, 2009; McCormick, 2008). Due to the measure-SLO misfit, practitioners gleaned limited or inaccurate data regarding student learning (Ewell, 2009). Since then, researchers and higher education practitioners have developed and adopted more sophisticated measures of student SLOs designed to measure specific outcomes such as critical thinking (Ewell, 2009).

Measuring what students know and what they gain may be sufficient for accountability. However, for those interested in improving how much students learn, being able to measure student learning is only the beginning. As pointed out by Fulcher, Good, Coleman, and Smith (2014), a pig does not get fatter just by weighing it. In other words, improving student learning requires more than measuring learning, it also requires
altering the learning environment.

While numerous interventions have been proposed for improving student learning, most improvement efforts can be categorized into one of four dimensions: curriculum/programming, teachers/teaching, extra-curricular learning opportunities, or students. Efforts to improve learning through curriculum or programming tend to revolve around changing program content, scaffolding, alignment, and resources. Educators have also attempted to increase student learning by changing teachers/teaching through faculty development. The underlying assumption of faculty development efforts seems to be that training teachers in content knowledge and pedagogy will improve teaching and thus increase student learning. Efforts to improve student learning may also focus on providing students with opportunities to learn outside of the classroom. These efforts may be represented by supporting learning resources such as math labs and writing centers or clubs relevant to student learning (e.g., math club). Educators also attempt to increase student learning by changing the way students interact with their learning environments, often described as student engagement (Kuh 2002; 2003). Overall, it seems educators may assume that well-reasoned changes in any one of these four areas might lead to students learning more than they did before.

The influence of each area on student learning is not independent of the other areas. Each piece works with the others to form a pathway between what students could learn and what they actually do. Consider a student enrolled in an algebra course. In this course, there is an ideal block of information and skills that the student should learn. We might hope that by the end of the course the student would have learned everything in the ideal block. However, in reality, the amount of information the student learns from that
block is dependent on each of the four factors: curriculum/programming, teachers/teaching, extra-curricular learning opportunities, and the student’s interaction with their learning environment. Figure 1 illustrates a hypothetical scenario relating these variables. Imagine that a teacher only covers 75% of the intended content and teaches it in an uninspired way. The block of ideal information has been attenuated to something less. Then, from that point, a student attends to social media in multiple classes rather than attending to the teacher and his peers. Furthermore, the student rarely takes advantage of supplemental resources and pertinent extra-curricular activities. Not surprisingly, the translation of the ideal information/skills to the students’ own knowledge and skills has been further diminished.

For the math student to learn the ideal information, he/she must first be exposed to it. In higher education, the most common conduit for this exposure is courses. However, for students to learn as much as they can from a course, the teacher must adequately expose students to the appropriate class content. This is the logic behind improving curriculum/programming and teachers/teaching. By improving either or both areas, educators attempt to improve the effectiveness with which the ideal information is presented to the student. However, to assume that improving teachers or programming automatically leads to increased student learning is to assume that educators are exclusively responsible for what students learn. This perspective assumes the student is a passive recipient of knowledge, only absorbing as much information as is presented to them.

Conversely, efforts to increase student learning through improving students’ interaction with their environment assume that students are active participants in their
learning. Student engagement is amongst the most prominent concepts discussed regarding students’ own influence on their learning. Engagement is often associated with connectedness, involvement, participation, motivation, investment, etc. (Baker, Schaufeli, Leiter, & Taris, 2008; Barger, Peltier, & Schultz, 2016; Dempsey, 2010; Domecq et al., 2014). Within higher education, engagement shares a similar connotation. Student engagement may be generally described as the degree to which students interact with learning opportunities, both within and outside of the classroom (Krause & Coates, 2008; Kuh, 2010; Kuh, Kinzie, Buckley, Bridges, & Hayek, 2007; Trowler, 2010). The underlying logic of engagement holds that the more students are involved in activities related to a subject the more they will learn about it.

Though researchers debate the specifics, they generally agree that students’ engagement is a major influence on their learning (Christenson, Reschly, & Wylie, 2012; Finn & Zimmer, 2012; McCormick, 2008; Trowler, 2010). Efforts to improve teaching and programming may provide students with greater exposure to the ideal information for a program. However, exposure alone is not enough for student learning. How much students learn is also dependent on how engaged they are with learning opportunities.

Consider two students, Irwin and Natalie. During their first year, both Irwin and Natalie enrolled in the same introductory psychology course. Recently the program offering the course had undergone restructuring to ensure all relevant content was covered. Additionally, the faculty member teaching the course had participated in professional development meant to improve the clarity and accessibility of her teaching. As a result, the course was well designed, well taught, and provided many opportunities for students to learn the content. However, despite the high quality of the course, Irwin
and Natalie differed greatly in how much they learned.

Natalie was deeply interested in the topics discussed in the psychology course. She completed all assignments, participated in class frequently, asked questions, and completed optional readings. In addition, Natalie joined a campus psychology club that provided her opportunities to discuss class content and be involved in research.

Irwin, while interested in psychology, was more focused on classes he knew would be directly relevant to his major. While he still completed assignments, he asked fewer questions and was less involved in class. Additionally, his other responsibilities prevented him from participating in any psychology-related extra-curricular activities.

At the end of the course, both Natalie and Irwin had learned a great deal. For each, the course provided all the relevant information and allowed them to benefit. However, because Natalie was more deeply engaged with the material and other learning opportunities, Natalie had learned far more than Irwin. The relationship between institutional factors and engagement in influencing student learning is visualized in Figure 2.

To add another dimension to the example, consider what may have been if Natalie and Irwin both enrolled in the course before it was improved. In this scenario, the introductory psychology course was designed to give students cursory knowledge of several important topics in psychology. However, the course had not been updated in years and was not aligned with other psychology courses in the program. Additionally, the class was taught by graduate students with little teaching experience. Given these environmental factors, students found the material unclear and uninteresting.

As before, Natalie and Irwin enrolled in the course. Natalie was deeply interested
in the psychology course, completed all assignments, participated in class frequently, asked questions, and completed optional readings. Natalie also joined the psychology club, providing her additional opportunities to learn. Irwin, though interested in psychology focused on major classes, completing only the necessary coursework and obligations for the class.

Like before, upon the conclusion of the course Natalie learned significantly more than Irwin. However, while Natalie and Irwin still differed from each other in how much they learned, their overall knowledge was significantly less than it would have been in the well taught and designed course.

Whereas Irwin left the well-prepared class with moderate knowledge of psychology, after the unprepared class, he left knowing little about the subject in the non-ideal scenario. Likewise, Natalie was fully engaged in both classes, however, the programming and teaching influenced how much she could learn from the course.

These examples illustrate two points. First, they demonstrate how improving student learning is not as simple as improving teaching. Rather, student learning is dependent upon the opportunities the institution provides students to learn as well as how much students take advantage of the opportunities they are given to learn. Second, differences between how much students learn is also dependent on their degree of engagement.

The role of student engagement in learning has made it a topic of interest to educators. Whether educators are conducting assessment for improvement or accountability, it is valuable to understand how students’ engagement affects their learning. The narrative that student engagement is a key factor in student learning seems
to be accepted throughout higher education (Christenson, Reschly, & Wylie, 2012; Kuh, 2001; 2009; McCormick, 2008; Trowler, 2010). Some have suggested that the relation between engagement and learning is such that increases in engagement can be a proxy for student learning (Kuh, 2009; McCormick, 2008). The link between engagement and learning is so accepted that institutions often use engagement to inform learning improvement efforts (Carini, Kuh, & Kline, 2006; Kuh, 2001; McCormick, 2008 Zilvinskis, Masseria, & Pike, 2017).

If engagement is to be used as a proxy for student learning, we might expect that research would consistently reveal moderate to strong positive relationships between student engagement and learning. Generally, research findings have supported the first premise; the more engaged students are the more they learn (Christenson, Reschly, & Wylie, 2012; Trowler, 2010). However, the research has not been as clear about the strength of the relationship between engagement and learning. Research has shown variability in the strength of relationships observed between engagement and learning, typically ranging from weak to moderate positive correlations. However, in some circumstances, engagement is used as a proxy for learning and as a basis for major institutional decisions. Such use of engagement measures implies a strong relationship.

There are several plausible explanations for the weak relationships commonly found in engagement research. One possibility is that students’ engagement may simply influence their learning less than other factors. If this were the case, engagement could still benefit educators but using engagement as the primary basis for decision making or as an indicator of learning would be questionable.

Other explanations include issues with the instruments, data, and measurement
techniques used in these studies. Before drawing conclusions about the relationship between engagement and learning, researchers must ensure that they are accurately measuring constructs. If engagement or learning outcomes are represented inaccurately, then the observed relationship will also be an inaccurate representation of reality. Several researchers have noted that the use of certain methodologies in the engagement-outcome literature may lead to inaccurate findings (Bowman, 2010; Gordon, Ludlum, & Hoey, 2008; LaNasa, Cabrera, & Trangsrud, 2009; Pascarella, 2001; Porter, Rumann, & Pontius, 2011; Porter, 2013). Nevertheless, warnings regarding these methods have generally gone unheeded.

Throughout this thesis, my focus will be on the relationship between engagement and student learning and how the methods used to study them may influence their observed relationship. To contextualize my research questions, the following chapter provides an overview of student engagement and learning outcomes, how each are conceptualized, measured, and their relationships with each other. Additionally, Chapter 2 describes how measurement issues may affect the relationship between engagement and learning outcomes.
CHAPTER TWO

Literature Review

Origin of Engagement Research

The foundational logic of engagement is that the amount of time, effort, and resources students devote to educationally meaningful activities is related to their learning (Graham, Tripp, Seawright, & Joeckel, 2007; Krause & Coates, 2008; Kuh, 2001; 2003; 2009; Kuh., Kinzie, Buckley, Bridges & Hayek, 2006; Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008; Trowler, 2010). Kuh, Cruce, Shoup, Kinzie, and Gonyea (2007) stated, “What students do during college counts more in terms of what they learn and whether they persist in college than who they are or even where they go to college” (p. 7). John Dewey, renowned philosopher, and educator, described similar ideas in his pedagogic creed (Dewey 1897; Graham, Tripp, Seawright, & Joeckel, 2007; Trowler, 2010). Though the concept of engagement is hardly novel, the systematic study of the relationships underlying engagement is a recent development in higher education (Alexander, 2000; Ewell, 2008; Finn & Zimmer, 2012; Kuh, 2009).

The exact genesis of modern engagement research is a point of disagreement amongst researchers (Kuh, 2009; Pike, Kuh, & McCormick, 2010; Trowler, 2010). Three scholars are typically credited: Ralph Tyler, Alexander Astin, and Robert Pace. The earliest of these is Ralph Tyler (1930) who studied the relationship between learning and time spent on a task (Kuh, 2009).

More commonly, engagement research is thought to have originated with Alexander Astin’s “Student Involvement Theory” (Astin, 1985; Ewell, 2008; Trowler, 2010). Indeed, Astin’s theory captures the foundational logic of engagement. Astin
stated, “Quite simply, student involvement refers to the amount of physical and psychological energy that the student devotes to the academic experience” (Astin, 1984, p. 518). To clarify his intended meaning of involvement Astin (1984) presented a list of similar verbs; these included “to partake,” “join in,” or “engage in.”

Astin (1984) suggested a positive relationship between college student involvement and personal and academic growth. He described student development as a function of the “quantity and quality of the physical and psychological energy that students invest in the college experience… such as absorption in academic work, participation in extra-curricular activities, and interaction with faculty or other institutional personnel” (Astin, 1984, p. 518). Astin’s theory held that student developmental outcomes not only depend upon content or teaching method, but also individual student behavior (Astin, 1984; 1993). The foundation of the engagement construct rests upon this principle. Because engagement and student involvement share such commonality, some researchers credit Astin as the father of modern engagement research (Astin, 1984; Pascarella & Seifert, 2010; Trowler, 2010). Indeed, there is little doubt that Astin’s research was instrumental in drawing greater attention to the relationships between student involvement and learning (Kuh, 2009).

Robert Pace’s research on quality of effort was also foundational in developing and popularizing the engagement construct (Kuh, 2001; 2009; McCormick, McClenny, 2012). In 1978 Pace began to study how students’ quality of effort predicts student learning and development (Pace, 1984). Pace (1984) spoke of educational experiences in two parts: products and processes. According to Pace (1984), a product refers to anything gained from one’s educational experience. Products of higher education may
include knowledge gained, new perspectives, and greater skills (Pace, 1984). Processes were a key part of Pace’s (1984) theory and refer to how students attain products. According to Pace (1984), both the quality of the “educational experience or process” as well as the product were necessary for determining educational quality (p.5)

Regardless of who is credited with the origin of modern engagement research, these authors all brought attention to the process that underlies student learning and development. Subsequently, engagement has become increasingly popular and researchers have devoted a great deal of effort to developing conceptualizations and measures of engagement.

**Conceptualizing Engagement**

As is common with many constructs, conceptualizations of engagement are plentiful, and settling on a primary definition has been a challenge for researchers (Appleton et al., 2008; Trowler, 2010). Perhaps the greatest difficulty has been the profusion of the “Jingle Jangle” fallacy in engagement research (Appleton et al., 2008; Reschly & Christenson, 2012). The Jingle-Jangle fallacy refers to either the utilization of two or more terms to describe the same construct (i.e., the “jingle”), or the use of one term to describe two or more constructs (i.e., the “jangle;” Finn & Zimmer, 2012). Such confusion of terminology can make navigating and understanding the engagement literature difficult. For example, terms such as student engagement, academic engagement, school engagement, and engagement with school all describe the same construct (Finn & Zimmer, 2012). Muddy definitions and conceptualizations have obfuscated the engagement literature. The issue has only been slightly ameliorated by researcher’s use of definitions and conceptualizations provided by experts such as George
Kuh, Ernst Pascarella, and Peter Ewell (Axelson & Flick, 2011).

While inconsistencies within the engagement research persist, common themes have emerged (Appleton, Christenson, & Furlong, 2008). First, engagement exists on a continuous scale ranging from fully engaged to disengaged (Appleton et al., 2008; Reschly & Christenson, 2012; Trowler, 2010). Second, engagement is plastic, capable of changing through intervention or over time (Fredricks, Blumenfeld, & Paris, 2004). Finally, engagement is best represented as a multi-faceted construct (Appleton et al., 2008; Fredricks et al., 2004; Trowler, 2010).

Fredricks, Blumenfeld, and Paris (2004) proposed a three-component model that has become one of the most popular conceptualizations of engagement. In a review of the literature, Fredricks et al. (2004) identified three components of engagement: behavioral, affective, and cognitive. Of these, behavioral engagement is the most common in research (Fredricks et al., 2004; Trowler, 2010). Fredricks et al. (2004) described behavioral engagement as educationally meaningful actions of students (e.g., studying, attending class, study abroad). Affective engagement refers to students’ feelings concerning their educational experience, including class content, teachers, or the institution. Finally, cognitive engagement refers to students’ investment in their educational experiences, including the amount of effort they are willing to put into comprehending and mastering the material (Fredricks et al., 2004).

While the model described by Fredricks et al. (2004) is prevalent, researchers have proposed alternative models of engagement. For example, Appleton, Christenson, Kim, and Reschly (2006) contend that engagement is best described using four components: academic, behavioral, cognitive, and psychological. Researchers have
proposed many similar models, the main differences between them is how they
categorize behavior.

While these models tend to focus on the student aspect of engagement, some
argue that such information is insufficient. Like Pace (1984), these researchers noted that
student learning is dependent upon both the student and institution (Axelson & Flick,
2011; Kuh, 2009; Quaye & Harper, 2014). In other words, engagement depends upon
both the students’ intrinsic will to engage as well as the opportunities provided by the
institution to be engaged (Pace, 1984; Trowler, 2010). Such logic seems obvious, and
student and institutional aspects of engagement have been accepted as critical
components in understanding student learning.

Up to this point, I have described research presenting engagement as a multi-
faceted construct, consisting of behavioral, cognitive, affective, and institutional
components. While researchers disagree how to best define engagement, the definition
provided by the NSSE is one of the most widely accepted. NSSE presents the following
information on its website:

Student engagement represents two critical features of collegiate quality.
The first is the amount of time and effort students put into their studies and
other educationally purposeful activities. The second is how the institution
deploys its resources and organizes the curriculum and other learning
opportunities to get students to participate in activities that decades of
research studies show are linked to student learning (NSSE: About NSSE,
n.d.).

Based upon decades of research, the NSSE defines engagement as a multi-faceted
construct composed of institutional and individual characteristics and behaviors (Kuh, 2001; 2003). Because of its alignment with previous research and my use of NSSE in this study, I have chosen to adopt NSSE’s definition of engagement. While conceptualizing and defining engagement are essential, researchers have also devoted considerable effort in developing measures of engagement such as the NSSE.

**Measuring Engagement**

Over the decades researchers have proposed many measures of engagement. Of these, the NSSE is by far the most prominent in higher education. Because the NSSE is also the main measure of engagement in this study, I devote the most time to it. The story of NSSE, however, is incomplete without a review of its predecessor, the College Student Engagement Questionnaire.

**College Student Engagement Questionnaire (CSEQ).** The CSEQ was developed by Robert Pace (1984), a scholar who researched students’ quality of effort (Kuh, 2009). Pace (1984) described education in two parts, the product or the outcome of education, and the process, or the way in which the product is attained (Pace, 1984). Pace (1984) argued that just as the quality of outcome or product differ, so does the quality of the processes. He thought that knowing how much time students put into a task was not sufficient to understand educational products (Pace, 1982; 1984). Rather, it is the quality of the practicing, studying, or other activities that provide rich information about the learning process (Pace, 1982; 1984). To help educators and researchers understand the role of quality of effort in learning, Pace (1982; 1984) developed and released the CSEQ.

The CSEQ was well used, with over 140 institutions participating when it was
discontinued in 2014. The most recent version of the CSEQ was released in 2007 and focused on measuring three aspects of the student experience: college activities, college environment, and student gains toward outcomes (CESQ: Content, 2007). Each of these three components provided key information for estimating the quality of students’ educational experiences and how much they learned. College activities, the first component of the CSEQ, included library experience, clubs, and organizations, art music and theater, experiences with faculty etc. The college environment component measured characteristics of the educational environment as well as scales concerning relationships between students and educators. Finally, student learning gains were measured by self-report on key outcomes such as science knowledge, intellectual skills, personal development, etc.

While it was not intended to measure engagement as it is described today, the CSEQ was a pioneer for measures of engagement. This is especially true for the NSSE. The NSSE developers drew heavily from the CSEQ, adopting much of its structure and over two-thirds of its items (Kuh, 2001; McCormick & McClenny, 2012).

**National Survey of Student Engagement.** While the CSEQ was well recognized, it was predominantly limited to educational researchers, with few institutions using it for improvement purposes (Kuh, 2001; 2009). Researchers recognized the need for a tool accessible to institutions, designed to be used for improvement and accountability purposes (Kuh, 2001; McCormick & McClenny, 2012; Pike, 2013). To this end, an initiative began. These efforts were hosted at the Indiana University Center for Postsecondary Research and involved several well-known researchers of educational quality such as Alexander Astin, Robert Pace, George Kuh, and Peter Ewell (Ewell,
Rather than reinvent the wheel, these researchers recognized the value of previous research and adopted much of the structure and many items from other measures, particularly the CSEQ. Like the CSEQ, the NSSE contains three main sections measuring college activities, college environment, and student gains towards outcomes (CSEQ: Content, 2007). The NSSE also contained many of the same items as other measures of engagement. However, the purposes of the NSSE are distinct from those of other measures of engagement (Kuh, 2009). Kuh (2001; 2009) stated three purposes of the NSSE: First, to serve as an accessible tool for measuring collegiate quality and to identify areas of improvement, second, to determine effective educational practices, and finally, to encourage the use of empirical measures of educational quality.

NSSE has certainly fulfilled its mission of creating an accessible measure of student engagement. Of the many measures of engagement, the National Survey of Student Engagement has become the most widely used and most influential with a total of 1642 institutions having taken part since its creation in 2000 (Kuh, 2009; NSSE: Participants, 2017).

Recall that engagement’s value is predicated on the assumption that more engaged students will learn more. Indeed, NSSE’s three goals are dependent upon its ability to predict desirable outcomes of higher education (Kuh, 2001; 2009; Pascarella & Seifert, 2010). Consistent with the general assumption of engagement leading to learning and to address NSSE’s three goals, the instrument is designed to measure student participation in effective educational practices (Kuh, 2001).

The NSSE’s organization currently includes four themes and eleven indicators of
effective educational practice, that is behaviors and institutional characteristics believed to be related to student learning outcomes (Table 1; Kuh, 2001; 2003; 2009; Pike, 2013; McCormick & McClenney, 2012). According to Kuh (2001), participation in activities measured by the NSSE often require substantial time and effort but lead to desirable outcomes. For example, the “Higher-Order Learning” theme is student-focused, and contains items such as “During your experience at JMU, how much has your coursework emphasized the following: Applying facts, theories, or methods to practical problems or new situations?” (NSSE: Engagement indicators, n.d.).

Finally, in addition to collecting student and institutional engagement data, NSSE measures SLOs. While NSSE’s focus is student engagement, SLOs are valuable to educators (McCormick & McClenney, 2012). Specifically, the NSSE measures self-reported learning gains on several outcomes, displayed in Table 2 (Zilvinskis, Masseria, McKinney, & Pike, 2015)

**NSSE validity.** No doubt researchers have spent considerable time developing the NSSE and thousands of institutions use the instrument. Regardless, the trustworthiness or “validity” of any instrument’s results, including the NSSE, should be examined. According to Messick (1990), validity is the process of evaluating the “appropriateness and adequacy” of a measure and its uses through theoretical argument and empirical evidence. While a comprehensive evaluation of the assumptions underlying uses of NSSE data is beyond the scope of this study, I will describe several key assumptions surrounding the use of NSSE data and relevant evidence.

First, the NSSE assumes a strong positive relationship between engagement and learning outcomes. As described earlier, many educators value engagement for its
relationship to desirable outcomes (Christenson, Reschly, & Wylie, 2012; Kuh, 2001; 2009; Trowler, 2010). Whether explicitly stated or not, engagement is seen as an indicator, if not a proxy, for outcomes when other outcomes measures are unavailable (Kuh, 2009). Thus, with the NSSE’s claim to provide information for institutional improvement, we must be sure that the relationship between engagement and learning outcomes is well supported (Pike, 1995).

Second, the NSSE assumes that it truly measures the engagement construct. If this were so, the NSSE should be related to other measures of engagement as well as the outcomes commonly associated with engagement (Kuh, 2001; 2009; Pascarella & Seifert, 2010).

The following sections explore the validity of the NSSE according to the areas of evidence set forth by the Standards for Educational and Psychological Testing: test content, response process, internal structure, and relations to other variables (American Educational Research Association, 1999). Unfortunately, limited research on the NSSE 2.0 prevents extensive analysis in the following sections. While the two versions are not identical, test content was not drastically altered in the update. Thus, where appropriate, evidence from the NSSE 1.0 supplements that of NSSE 2.0

**Test Content.** Substantive validity refers to how well the instrument theoretically relates to the construct it is intended to measure (Benson, 1998). Evidence of this sort is well documented; manifest by its development by content experts, reliance on decades of engagement research, connection to conceptually similar instruments, and consideration of measurement concerns.

As previously described, NSSE development team was comprised of well-known
educational researchers. Their intent was to develop a measure of engagement that would inform institutions about students’ participation in practices tied to desirable learning outcomes (Kuh, 2001; NSSE: my origin and potential, n.d.). As such, the development of the NSSE was anchored in both research and professional experience.

As engagement practices are linked theoretically to student learning outcomes, the validity of the NSSE is dependent upon its ability to predict learning outcomes (Astin, 1991; Krause & Coates, 2008, p. 493; Pascarella & Terenzini, 1991). In an effort to develop engagement items that would capture this relationship, the research team drew upon research and theory suggesting ties between specific practices and outcomes (Kuh, 2009; NSSE: my origin and potential, 2001). Chickering and Gammons’ (1987) “Seven principles of good practice in undergraduate education” was one of these (Kuh, 2009). The seven principles are as follows,

1. Encouraging contact between students and faculty
2. Encouraging reciprocity and cooperation amongst students
3. Encouraging active learning techniques
4. Giving students prompt feedback
5. Emphasizing time on task
6. Communicating high expectations
7. Respecting diverse talents and ways of learning

(Chickering & Gamson, 1987, pp.1)

Evident in Chickering and Gammons’ (1987) principles, the behaviors associated with engagement comprise a broad range of practices. Likewise, the NSSE was designed to measure engagement in a wide range of practices, grouping similar practices into
themes (Kuh, 2001; 2009; NSSE: our origins and potential, 2001). Many such practices, known to be tied to outcomes, were also measured by the CSEQ and others predating NSSE (Kuh, 2009; Kuh, Pace, Vesper, 1990; NSSE: our origins and potential, 2001; Pace, 1984; 1995). The NSSE developers recognized the strength of these instruments, incorporating pieces of them after careful review. Many items were taken from the CSEQ and incorporated into the NSSE (Kuh, 2009; NSSE: our origins and potential, 2001). Thus, upon creation, the NSSE was a compilation of items expected to predict outcomes.

However, the NSSE has been revised since its original creation. The original NSSE only described engagement at the broadest levels. To provide a more detailed description of student engagement, an updated version of the NSSE was released in 2013 (McCormick et al., 2013; Pike, 2013). This version included a restructuring of the five original themes into four themes and ten specific indicators (Table 1; Pike, 2013). Through this revision the focus on the NSSE’s relationship to learning outcomes remained unchanged. Many of the original items are still present in the NSSE 2.0, though some were reworded for clarity (McCormick et al., 2013; NSSE: Benchmarks to engagement, 2014; NSSE: Survey instrument, 2017).

Unfortunately, there is very little independent research that uses the NSSE 2.0, likely due to its newness and the lag time between research and publication. While there is little evidence supporting the validity of the NSSE 2.0, it seems reasonable that changes would only improve the validity of the NSSE. Regarding substantive validity, the NSSE revisions were based upon research indicating a need to better represent specific types of engagement.
In addition to constructing a solid theoretical foundation, the developers of the NSSE were cognizant of measurement issues, particularly the use of self-reported measures (Kuh, 2001; 2002; 2004; McCormick et al., 2013). As discussed in earlier sections, it is common for researchers to be distrustful of self-reported behavioral measures. With this in mind, researchers selected or developed NSSE items using five guidelines for self-reported validity established by prior research (Kuh, 2001; 2002; 2004). These guidelines suggest that self-reports are valid if the information requested is: (1) known by the student, (2) not embarrassing or threatening, (3) worded clearly, (4) deserving of serious thought, and (5) referencing recent activity (Kuh, 2002; Pace, 1984).

However, even those championing a self-reported approach concede that responses may deviate somewhat from the true response (Kuh, 2002; Pace, 1984; Pike, 1999). Thus, while engagement, as measured by the NSSE, may not be exactly representative of students’ actual engagement, evidence suggests that they may be relatively similar (Pike, 1999). Overall, these data suggest a strong theoretical foundation for the NSSE, representative of the engagement construct.

**Response Process.** Evidence based on response process refers to how participants understand and interpret the instrument, its items, instructions, and purpose. This type of evidence is often collected by asking participants to do things such as reflect on their response process, talk aloud while responding to questions, and interpret the meaning of items. These activities are meant to give insight as to whether participants response processes are in line with the purposes of the instrument. There has been little research looking at the response processes of students taking the NSSE. The primary source of information about NSSE response process are provided in NSSE’s online psychometric
portfolio. In the process of revising the NSSE data from several focus groups and
cognitive interviews were collected. As part of the cognitive interviews students were
asked to answer questions from this survey and describe their interpretation of the
questions asked. Additionally, students participating in these interviews were asked to
think aloud as they responded to the questions on the NSSE. In all cases participants
were asked to identify any confusing questions. Generally speaking, the focus groups
and cognitive interviews conducted by NSSE suggested that students of varying
backgrounds interpreted the NSSE items similarly and in a way that was consistent with
the intent of the instrument developers. Despite the lack of independent research, what
has been conducted suggests that the NSSE is valid in terms of the response processes of
those taking the NSSE.

**Internal Structure.** Structural validity refers to the psychometric properties of
data related to an instrument, such as reliability, factor structure, and inter-item
relationships (American Educational Research Association, 1999). Evidence of this sort
is plentiful but mixed in its findings on the NSSE. Recall that the NSSE Themes are
items grouped together based on statistical and theoretical evidence (Kuh, 2009;
McCormick & McClenny, 2012; Pike, Kuh, McCormick, Ethington, & Smart, 2011b).
Using principal components analyses (PCA) the NSSE developers clustered the questions
according to a structure of commonality (Kuh, 2009).

Most NSSE validity research has been conducted on the original structure.
However, the development NSSE 2.0 introduced a new way to structure the items, as
well as clarified item wording (NSSE: Benchmarks to engagement, 2014; Pike, 2013).
Such changes have the capability to change the structural validity of an instrument.
Because of this, I will briefly review the structural validity of the NSSE 1.0 and 2.0. It should be noted here that many researchers have combined items in numerous ways beyond what was initially proposed by NSSE. Speaking of all possible combinations is beyond the scope of this thesis. Therefore, I will primarily concentrate on the structures proposed by NSSE.

**Structure and reliability.** Across the board, evidence of NSSE’s structural validity has been mixed (Campbell & Cabrera, 2011; LaNasa et al., 2009; McCormick & McClenny, 2012; Pike, 2013). The discussion surrounding NSSE’s structural validity consists of two parts: philosophical debate and psychometric perspectives. Studies exploring the psychometric properties of the NSSE use a range of methodologies, each of which have their own philosophical assumptions. The most prominent of these are principle component analysis (PCA) used with observable variables, principle factor analysis (PFA), and confirmatory factor analysis (CFA) used with latent variables.

When the NSSE was first constructed, researchers used a blend of theory and PCS to determine how items should be structured into themes*(NSSE: my origins and potential, 2001). Their original findings, as well as subsequent investigations, provided strong structural validity of the evidence for NSSE’s structure. However, as independent researchers began to conduct their own analyses, many found that the NSSE’s structure did not hold in their sample (Campbell & Cabrea, 2011; Gordon et al., 2008; LaNasa et al., 2009; Lutz & Culver, 2010; Porter, 2011; Porter et al., 2011; Webber, Krylow, & Zhang, 2013).

In response to these studies, proponents of the NSSE defended its structure, explaining these findings as a result of improper methodology (McCormick &
McClenney, 2012; Pike, 2013). While the NSSE was originally designed using PCA, many researchers have relied on EFA and CFA to explore the NSSE’s structure. The issue with EFA and CFA is they assume the NSSE measures latent constructs (McCormick & McClenney, 2012; Pike, 2013). However, some have noted that the NSSE benchmarks were never intended to represent latent constructs, but rather broad categories of educational practice (Kuh, 2001; McCormick & McClenney, 2012; Pike, 2013). These claims seem inconsistent with the language used to describe the NSSE. Swerdzewski, Miller, and Mitchell (2007) noted that NSSE literature often refers to themes as factors, and suggests that they measure latent characteristics. This may have contributed to the confusion surrounding NSSE’s structure.

Because of the confusion surrounding the philosophical underpinnings of the NSSE, there is little independent research exploring its structural validity using analyses conducted at the institutional level. However, research conducted by NSSE and its affiliates have generally supported its structure (Kuh, 2002; NSSE: Validity, n.d.).

NSSE scores have generally shown acceptable reliability (Kuh, 2001a: NSSE: Reliability, n.d.; Pascarella et al., 2009; Pike, 2006). More specifically, multiple sources describe reasonable consistency of scores (α > .70) and strong correlations between NSSE benchmark scores over time (NSSE: Reliability, n.d). However, some research has indicated reliability that is less than satisfactory (Gordon et al., 2008; Porter, 2011). Nevertheless, this reliability was not drastically different than that reported by NSSE (Pike et al., 2011).

Most of the research on the reliability and validity of the NSSE was conducted on its original form. While the NSSE maintains many of its original pieces, its structure has
changed, and thus the evidence relating to the original NSSE may no longer be applicable. What little research that has been done on the structure of the NSSE appears promising (Zilvinskis et al., 2015).

Recall that when the NSSE was revised the original five themes were divided into a set of ten indicators. According to research conducted on the original NSSE, models with more factors tended to provide a better structural fit (Porter, 2011; Tendhar, Culer, & Burge, 2013). Additionally, an increase of factors also means that information provided by the NSSE may be more detailed and provide a better prediction of specific outcomes (Pike, 2006a; 2006b). According to the limited research, this claim seems to be supported.

A study by Miller, Sarraf, Dumford, and Rocconi (2013) investigated the factor structure of the ten engagement indicators. Using EFA, they found that ten components matched the ten EIs predicting 60% of the variance. Additionally, CFA was used to explore the structure of the indicator themes. Overall, their results suggested “adequate” to “very good” model fit for the categorization of the indicators (Miller, Sarraf, Dumford, & Rocconi, 2013). However, these researchers used the techniques criticized for use in studying the previous version of the NSSE. Because of this, it is not clear how much structural evidence these findings actually provide for the validity of NSSE scores.

While it is clear that more research should be conducted on NSSE’s structural validity, it continues to be the most common measure of student engagement. To further explore the NSSEs use in measuring engagement in higher education I now turn to a brief discussion of the external validity of the NSSE.

*Relations to Other Variables.* In addition to sound theory and structure, measures
should be consistently related to other theoretically related constructs. (American Educational Research Association, 1999; Cronbach & Meehl, 1955; Campbell & Fiske, 1959; Messik, 1995). In terms of the NSSE, it should relate to other measures of engagement, and learning outcomes.

The relationship of NSSE data with learning outcomes is especially important in establishing the valid use of the NSSE as an indicator of learning outcomes. Campbell and Cabrera (2011) stated, “If the NSSE themes* are a valid measure of student engagement, they should be predictive of student learning across a variety of institutional types and student populations (i.e., have predictive validity)” (pp.80).

Supporting NSSE’s external validity, many studies have found relationships between NSSE scores and learning outcomes. Additionally, the updated version of the NSSE may better predict learning outcomes than the original. Zilvinskis, Masseria, and Pike (2017) compared the original NSSE and the new version’s ability to predict learning outcomes using canonical correlation. Their findings suggested that both versions of the NSSE were strongly related to self-reported learning gains (NSSE 1.0: $R_c=.84$, $R_r=.62$; NSSE 2.0: $R_c=.85$, $R_r=.75$. Additionally, they found that the NSSE 2.0 scales could predict learning outcomes more precisely than those in the NSSE 1.0. From these findings they concluded, “From a practical standpoint, the engagement indicators provided in the new NSSE survey appear to be more useful than previous engagement measures in identifying institutional actions that can enhance certain types of learning outcomes.”

The relationship between engagement and learning outcomes has been frequently studied and will be discussed in greater detail later as it is central to this thesis. Before
exploring the engagement-outcome relationships more deeply, I will discuss a few measurement issues of the NSSE followed by a brief introduction to learning outcomes.

**Additional Measurement and Design Issues.** Two measurement issues commonly arise regarding the NSSE: volunteer sampling and use of self-reported measures. Because participation in the NSSE is voluntary, some have voiced concern that the NSSE samples may not be representative of the target student population. The primary concern is that students with diverse levels of engagement may not have the same probability of responding. If this were the case, it could mean that the NSSE only collects data from the most engaged of students. Research looking into this possibility has not found any meaningful difference between students who respond to the NSSE and those who do not (NSSE: Other quality indicators, n.d.; Sarraf, 2005). The differences that have been found have been notably small and inconsistent (Kuh, 2003; NSSE: Other quality indicators, n.d.).

A second common concern is NSSE’s use of students' self-reported engagement behaviors. Some have questioned whether students’ estimates of how frequently they engage in a behavior is representative of reality. Those developing the NSSE were aware of this possibility. To increase accuracy, they followed guidelines research had suggested for accurate self-reporting (Kuh, 2002; 2009).

Despite these efforts, some have suggested that students’ self-reported behaviors may still be inaccurate (Gonyea, 2005; Pascarella et al., 2010; Porter, 2001; Porter, 2011). In a review of the literature, Porter (2011) found that students have difficulty accurately describing their behavior over time and may describe themselves in a more positive light than is accurate. In a similar vein, Pike (1999) found that students were
prone to overestimate their participation in activities as well as their learning gains.

While acknowledging the existence of these issues, some have noted that actual behaviors and self-reported behaviors are still correlated (Gawronski, LeBel, & Peters, 2007; Gonyea, 2005). Thus, self-reported engagement may still provide some estimation of students' actual behavior and is easier to collect than direct measures (Fredricks & McColskey, 2012; McCormick & McClennen, 2012; Pike, 1995; 1996).

The NSSE also includes self-reported measures of gains on SLOs. Discussion about the accuracy of such measures is an ongoing debate and is central to the research contained in this study. Before discussing this issue, I will return to a topic broached in the external validity section: learning outcomes and their role in educational research.

**Student Learning**

Student learning outcomes (SLOs) are the products of educational experiences, that is, what students know, think, or do because of their educational experience (Dugan & Hernon, 2002; Harden, 2002; Hussey & Smith, 2003; Melton, 1996; Spady, 1988). As discussed in Chapter 1, this kind of information is particularly valuable in higher education. It can be used for accountability, to demonstrate the value of higher education, or for improvement purposes (Ewell, 2005; Gonyea, 2005; Pike, 1995). One of the most valuable uses of learning outcomes is providing evidence of student learning.

It is easy to see how evidence of student learning would be valuable to educators and policymakers. Student learning, represented by gains in learning outcomes and competency, demonstrates the value of educational experiences (Ewell, 2005; Pike, 1995). Additionally, knowing what students gain from an experience can help educators determine what changes should be made to improve student learning. Comparisons of
gains in learning may also help educators determine if an intervention was effective (Ewell, 2005; Fulcher et al., 2014). Such outcomes are measured at many levels, most commonly institutional, programmatic, or course level (Ewell, 2005).

According to Nusche (2008), outcomes may be organized into both cognitive and non-cognitive types. Measures of cognitive outcomes are the most prevalent and typically include content knowledge, and development of physical or intellectual skills (Gonyea, 2005; Nusche, 2008). Non-cognitive outcomes are less common and typically describe values, beliefs, or attitudes of students (Ewell, 2005; Nusche, 2008).

**Measuring outcomes.** The value of learning outcomes in higher education has inspired the creation of many instruments. Learning outcome instruments vary widely, some focusing on general education and soft skills outcomes while others emphasize specific content knowledge. One example is the HEIghten suite by ETS. The HEIghten suite measures three learning outcomes: critical thinking, quantitative literacy, and written communication (ETS: HEIghten, 2017). Other measures include ACT’s Collegiate Assessment of Academic Proficiency critical thinking test and the Collegiate Learning Assessment Plus.

In addition to measuring engagement, the NSSE includes a brief set of questions measuring learning outcomes (NSSE: Survey instruments, 2017). While the outcome section of the NSSE is brief, it attempts to measure learning gains on more outcomes than most instruments devoted to measuring outcomes (recall, HEIghten only measures three outcomes). It may seem surprising that a tool devoted to measuring outcomes relates to fewer outcomes than one devoted to engagement. Why is this the case? The difference is a matter of methodology. The HEIghten uses direct measures of student learning while
the NSSE relies on indirect measures.

**Direct /Indirect measures.** Often, the decision to use direct or indirect measures relates to type of construct and convenience (Nusche, 2008; Pike, 1996). Recall that non-cognitive outcomes often include unobservable topics such as emotion, attitudes, or other similar traits. For these variables, the best option may be to use indirect measures such as self-reports (Gonyea, 2005; Nusche, 2008). Conversely, cognitive outcomes (such as students’ knowledge of mathematics) lend themselves more easily to observation through direct measures (Nushe, 2008).

Direct measures are often considered the gold standard of measurement (Gonyea, 2005). Most people would more readily trust direct observation of a behavior over someone’s self-report of the behavior. For example, common sense would indicate that evaluating students’ writing by reviewing their essays would be more accurate than their self-report of writing proficiency. Indeed, it is often assumed that direct measures are less subjective than self-reports.

While direct measures of cognitive outcomes are likely more accurate than self-reports, self-reported learning gains (SRLG) are used more frequently than direct measures when reporting learning gains (Pike, 1993; 1995; Price & Randall, 2008; Sitzman, Ely, Brown, & Bauer, 2010). The preference for self-reports may be attributed to the convenience of collecting self-reported data (Pike, 1996). Self-reports are often cheaper, require less time to complete, and are less time-consuming to develop. This convenience makes self-reported gains particularly attractive as potential proxies for direct measures of gains.

**Self-Reported Learning Gains as Proxies for Direct Measures of Learning Gains**

Researchers have questioned the validity of using SRLG data as proxies for Direct
Measures of Learning Gains (DMLG) data (Bowman, 2010; Gonyea, 2005; Kuh, 2001; Pace, 1984; Pike, 1995; 1996; Porter 2011; 2012; 2013). To be accurate, self-estimated learning requires students to accurately recollect and compare their ability/knowledge at two time points. Several researchers have questioned if students do, or are capable of, accurately estimating these gains (Bowman, 2010; Carrell & Willmington, 1996; Gonyea, 2005). Pike (1995; 1996) noted that if direct and self-reported questions measure the same construct they should be highly correlated, and therefore valid proxies for each other (Pike, 1995; 1996). With few exceptions, research exploring the relationship between SRLG and DMLG has reported weak to moderate relationships between self-reported and direct measures of the outcome (Bowman, 2010; Carrell & Willmington, 1996; Gonyea, 2005; Pike, 1995; 1996; Pohlmann & Beggs, 1974; Price & Randall, 2008; Sitzmann et al., 2010). Following Pike’s (1995; 1996) logic, these findings would suggest that while SRLG and DMLG may focus on the same topic, each measures a different construct (Carrell & Willmington, 1996; Pike, 1996). In other words, DMLG of mathematical ability may represent students’ skill, SRLG of mathematical ability measure students’ perception of their skill. This misalignment may explain the weak to moderate relationships between self-report and direct measures.

Other factors may influence the strength of relationships between SRLG and DMLG. As with self-reported engagement, research has found that students typically overestimate their knowledge or skills (Luce & Kirnan, 2016; Pike, 1999; Porter, 2011). What’s more, low ability students tend to overestimate their ability more than high ability students do, a phenomenon known as the Dunning-Kruger Effect (Cole & Gonyea, 2010; Kruger & Dunning, 1999; Kuncel, Crede, & Thomas, 2005; Luce & Kirnan, 2016).
Combined, these findings suggest that self-reported measures may not accurately represent students’ ability or learning (Gonyea, 2005; Pike, 1995; 1996; Price & Randall, 2008; Sitzmann et al., 2010).

Pike (1993; 1995; 1996) suggested that weak relationships might be due to misalignment of measurement content; when self-report and direct measures ask about similar content they are more strongly correlated. Additionally, researchers have noted that relationships might be attenuated by the difference in scope of measurement between methods (Astin, 1993; Pace, 2005; Pike, 1995). Whereas self-reports are typically broad, direct measures tend to be more focused on specific knowledge or abilities (Pike, 1995). Support for this explanation has been found in several studies (Dumont & Troelstrup, 1984; Pike, 1995).

In summary, self-reported measures may be appropriate for some types of content; however, their use as proxies for direct measures of learning outcome gains is not supported by research (Bowman, 2010; Carrell & Willmington, 1996; Gonyea, 2005; Pike, 1995; 1996; Pohlmann & Beggs, 1974; Price & Randall, 2008; Sitzmann et al., 2010). Self-reports may be less representative of students’ ability than a well-validated direct measure of ability (Gonyea, 2005; Pike, 1995; 1996; Price & Randall, 2008; Sitzmann et al., 2010). Therefore, researchers should avoid using SRLGs as proxies for DMLGs and practice caution when interpreting SRLGs (Bowman, 2010; Gonyea, 2005; Sitzmann et al., 2010).

**Engagement and its Relationship to Learning Outcomes**

Considering the close theoretical relationship between student engagement and learning, it is not unreasonable to think that engaged students may attain a greater level of skill/knowledge than their disengaged counterparts (Campbell & Cabrera, 2011;
Campbell & Fiske, 1959; Carini, Kuh, & Kline, 2006; Cronbach & Meehl, 1955; LaNasa et al., 2009; Messik, 1995; Zilvinskis et al., 2015). At the broadest level, a relationship between engagement practices and learning outcomes seems a matter of common sense.

However, specific learning outcomes and engagement practices can take many forms. Exploring relationships between these specific components is particularly important for higher education. First, evidence of specific engagement-outcome relationships confirms the value higher education places on student engagement. Second, understanding this specific relationship can provide information about what engagement practices the institution may focus on to improve specific outcomes.

This section explores research investigating the relationship between engagement and specific outcomes. Additionally, this section includes evidence of the predictive validity of NSSE scores. To do so, research findings are separated into two parts: relationships found using the NSSE and those found using other measures of engagement. Separating results in this way demonstrates the predictive validity of NSSE data by allowing for easy comparison of similarities and differences in engagement-outcome relationships.

**GPA.** Of all the learning-related variables studied in relation to engagement, GPA is the most common. GPA seems to offer two distinct advantages. First, it is an accessible and inexpensive indicator of academic achievement. A second advantage is its ability to account for academic achievement across time and learning subjects. Because of this, GPA is a macro indicator of student achievement, and theoretically, learning. Thus, GPA is frequently used to explore the relationship between engagement and academic achievement/learning at the broadest level.
Engagement-learning research consistently shows weak to moderate relationships between engagement and GPA regardless of whether engagement was measured using the NSSE (Carini, Kuh, & Klein, 2006; Fuller, Wilson, & Tobin, 2011; Gordon et al., 2008; Kuh, 2003; Kuh et al., 2008; Kuh et al., 2006; LaNasa, Olson, & Alleman, 2007), or some other measure (Astin & Sax, 1998; Gonyea, 2006; Kuh et al., 2006; Pike, 1999; Svanum & Bigatti, 2009; Trowler, 2010).

The strength and consistency of these results are particularly important to the discussion of engagement-outcomes relationships as well as NSSE predictive validity. First, the presence and similarity of relationships across measures of engagement speak to the predictive validity of the NSSE. While weak, the similarity between the NSSE and non-NSSE measures in their relationship to learning outcomes suggests that the NSSE is as effective as other measures of engagement.

Second, while statistically significant, findings do not support the assumption that engagement and outcomes are strongly related. Recall that with large sample sizes, even weak relationships can be statistically significant. Because many engagement studies have a large sample size it is unsurprising that engagement is weakly and significantly tied to GPA (Gordon et al., 2008; Fuller et al., 2011; Korzekwa, 2010). One reason for the weak relationships may be that GPA is influenced by many factors unrelated to student learning (e.g., differential grading policies, attendance, etc.). Additionally, many of these studies tested the relationship between specific engagement practices and GPA. Because GPA is so broad, it seems unlikely that any single engagement subscale would be strongly related to GPA.

**Engagement-Knowledge/Ability.** Whereas GPA measures achievement at a
macro level, other learning-related variables provide further insights into engagement-learning outcome relationships. Engagement as measured by both NSSE and non-NSSE instruments has been linked to performance on many measures of learning outcomes, including the Collegiate Assessment of Academic Proficiency (CAAP), RAND tests, and parts of the GRE (Benjamin & Hersh, 2002; Carini et al., 2006; Ewell, 2002; Fuller et al., 2011; Kuh, 2002; Pascarella et al., 2009; 2010).

Relatedly, much research explores the relationship between engagement and liberal arts/general education outcomes (e.g., reading/writing, critical thinking, mathematical ability etc.). Frequently, research finds that data from non-NSSE measures of engagement is related to:

- Writing ability (Gonyea, 2006; Kuh, Pace, & Vesper, 1997)
- Critical thinking (Carini et al., 2006; Gonyea, 2006; Kuh, Pace, & Vesper, 1997; Terenzini, Pascarella & Blimling, 1996; Pace, 1984; Pike, 1999; Terenzini & Pascarella, 1995)
- Quantitative ability (Gonyea, 2006; Kuh, Pace & Vesper, 1997)

Studies using the NSSE have reported similar relationships,

- Writing ability (Blaich & Wise, 2011; Pascarella, Seifert & Blaich, 2009)
- Critical thinking (Carini, Kuh & Kline, 2006; Pascarella et al., 2009)
- Quantitative ability (Blaich & Wise, 2011; Pascarella et al., 2009)

It is clear that using measures of specific outcomes provides greater detail about the engagement-outcome relationships than does GPA. While these findings are more detailed than the engagement-GPA relationship, the strength of the relationships are not drastically different than for GPA. Recall, engagement’s appeal is in its relationship with
student learning. While the relationships described by these studies show a relationship between engagement and knowledge, they do not tell us whether engagement is related to changes in student ability or knowledge. As discussed earlier, outcomes are often only measured at one time point and do not capture changes in the outcome over time. Such designs are common in engagement-outcome research. (Fuller et al., 2011; Pascarella et al., 2009; 2010).

**Engagement-SRLG.** While one-point-in-time data has value, it provides no information concerning how much students have learned. For example, students’ final exam math test scores only represent their current mathematical knowledge. To determine how much they have learned, researchers need to use a longitudinal design, measuring students’ mathematical ability before and after the course. However, collecting data from two timepoints is often difficult. One way researchers have approached this problem is by asking students to estimate how much they have learned regarding a specific outcome (SRLG). By using SRLG, researchers can collect information about longitudinal gains using one time point by asking students to estimate their learning. This has been the most common approach in research relating engagement to learning (Pascarella et al., 2010; Pike et al., 2011; Porter, 2011; Umbach & Wawrzynski, 2005; Zhou & Kuh, 2004).

Research relating non-NSSE engagement data to SRLG outcomes such as critical thinking, intellectual ability, reading/writing, or math, have typically found weak to moderate relationships (Carini et al., 2006; Kuh, Pace, & Vesper, 1997; Loes, Pascarella, & Umbach, 2012; Lundberg & Schreiner, 2004; Pace, 1982; Pascarella et al., 2014; Pike, 1999; Terenzini, et al., 1996; Terenzini, Springer, Pascarella, & Nora, 1995; Terenzini,
Likewise, researchers have also tied students NSSE scores to SRLG in critical thinking, intellectual ability, reading/writing, and math (Blaich & Wise, 2011; Carini et al., 2006; Fuller et al., 2011; Gonyea, 2006; Ishitani & McKitrick, 2010; Kuh et al., 2006; Kuh et al., 2008; Laird, Shoup, Kuh, & Schwarz, 2008; Pike, 2006; Pike & Kuh, 2005; Pike, Kuh, & McCormick, 2011a; Umbach & Wawrzynski, 2005; Zhou & Kuh, 2004).

**Engagement-DMLG.** Being the most convenient measure of student learning, it is not surprising that SRLGs are prevalent in engagement-outcome research (Loes et al., 2012; Pascarella et al., 2009; 2010; Pascarella et al., 2014). While SRLG provide some indication of student learning, some argue that student learning should be studied using longitudinal DMLG (Porter, 2013; Pascarella et al., 2009; 2010). Directly measuring student learning gains involves the collection of student outcome data from at least two time points. Using this data student ability/knowledge is compared across two time points with the difference considered to represent change in knowledge (i.e., learning gain). Currently, there is little-published research which uses this methodology to study engagement and student learning (Pascarella et al., 2009; 2010).

In my review I found two studies including examples of DMLG in engagement-outcome research. Both studies were conducted by Pascarella et al. (2009; 2010), in conjunction with the Wabash National Study of Liberal Arts Education (WNSLAE). According to Pascarella et al., (2009), “The WNSLAE is a longitudinal investigation of the institutional experiences that enhance growth in important educational outcomes (p. 7).” Both studies used institutional level data to explore the use of students’ NSSE data to predict learning gains in liberal arts outcomes. In both studies, researchers found weak to
moderate relationships between NSSE engagement data and gains in reasoning, problem-solving, inquiry, leadership, and cultural awareness (Pascarella et al., 2009; 2010). For example Pascarella et al., (2010) found significant moderate correlations between the CAAP Critical Thinking Test, level of academic challenge \((r=0.43)\) and the enriching educational experiences scale \((r=0.44)\). Additionally, Pascarella et al., found that critical thinking was moderately related to NSSE’s Higher order learning scale \((r=0.45)\). While these findings appear promising, further research using DMLG is needed. Concerns about methodological issues persist in NSSE-outcomes research; concerns which may draw the accuracy of the conclusions of engagement research into question. In the following sections, I will describe some of these concerns, their potential impact, and potential solutions.

**Methodological Concerns of Engagement-Outcome Research**

Recall that the premise of engagement is that the more time and effort students put into their education the more they should learn. Intuitively, this suggests a strong relationship between engagement and learning outcomes. However, research has found varying relationships between engagement and learning ranging from weak to moderate positive relationships.

As was mentioned earlier, there are many plausible explanations for the weak to moderate engagement-learning positive correlations. The simplest explanation is that engagement is not as strongly related to learning as typically thought. Student learning is influenced by several factors outside of engagement, limiting engagement’s ability to predict learning. Carini et al. (2006) stated, “…Learning outcomes stem from a variety of sources of which student engagement is only one” (Carini et al., 2006, pp.19).
However, before making these conclusions it is important to be sure that the relationships represent reality and are not an artifact of methodology. In other words, it is important to determine if the relationship affected by systematic error or if it is due to trait-variance.

Another explanation for the weak relationships may lie in the way our constructs are measured. If there is robust evidence supporting the inferences made from the measure of a construct, we would expect that the measure would accurately represent the construct it is supposed to measure. For example, based on our review of the NSSE’s validity, we would expect that it should accurately represent students’ perceptions of their general engagement. If the evidence supporting the inferences made from the measure is weak, the measure may be addressing a construct other than what was intended.

When exploring the relationships between two constructs if there is weak evidence supporting the inferences from either of the measures we cannot be confident that the relationship we observed is representative of the true relationship between our constructs. Thus, when considering the weak relationship between engagement and student learning it may be that their true relationship is strong but is not well represented by the measures being used. For example, often research examines how students’ engagement is tied to their learning in a specific area (e.g., quantitative literacy). If a researcher is interested in quantitative literacy and its relation to engagement, but uses an instrument measuring engagement across several study topics, the observed relationship will be that of student engagement to all their fields of study and their quantitative literacy. In this case we might expect to see a weak relationship, as the engagement measures reflects English, cultural studies, and others topics that are not be theoretically linked to the development of quantitative literacy. Thus, the observed relationships may
be weak due to a specificity mismatch.

In addition to specificity, whether an instrument directly or indirectly measures a construct may influence the claims that can be made about the relationships. For example, comparing student learning gains via longitudinal direct measure designs may be more accurate than self-report.

In the final portions of this review, I describe two aspects of measurement that may influence the results of engagement-outcome research. These are (1) if learning gains are measured via direct or indirect measures, and (2) the alignment of measures.

**Direct and Indirect Measures of Learning Gains.** As I have described, SRLGs – types of indirect measures - are the most common measure of student learning in engagement-outcome research. Whether the research is focused on exploring relationships between engagement and how students learn, or the predictive validity of NSSE data, most studies rely on SRLG (Blaich & Wise, 2011; Carini et al., 2006; Gonyea, 2006; Kuh et al., 1997; Kuh et al., 2006; Kuh et al., 2008; Laird, et al., 2008; Loes et al., 2012; Pace, 1982; Pike, 1999; 2006; Pike et al., 2011a; Terenzini & Pascarella, 1995; Terenzini, et al., 1996; Zhou & Kuh, 2004). As I have briefly discussed this may be problematic for engagement-outcome research and those who use it for practical application.

To understand why this may be problematic recall the measurement discussion from the outcomes section of this chapter. Several researchers have warned of the limitations of using SRLG with the NSSE (Gordon et al., 2008; Gonyea, 2005; Kuh, 2001; LaNasa et al., 2009; Pascarella et al., 2008; 2010; Porter, 2011; 2012). Pascarella (2001) noted that SRLGs are problematic because they provide no indication of student
knowledge upon entering college. They note,

When researchers do not have a precollege measure of an individual student’s receptiveness to educational experiences, it is difficult—if not impossible—to distinguish between how much of that student’s self-reported “gain” on some outcome is due to the added value of college and how much is simply due to his or her disproportionate openness and receptivity to the college experience (Pascarella et al., 2009 pg. 6)

In addition to being unable to identify the factors contributing to student learning, SRLGs are also subject to several errors. Students’ estimation of their ability and learning have both been shown to be subject to halo-effects and social desirability (Bowman, 2010; Kuh, 2001; Pike, 1999; Porter, 2010; 2012; Zilvinskis et al., 2017).

These findings suggest that students’ estimates of their own learning may not be representative of how much they have learned. Additionally, studies comparing students’ ability to estimate their learning against direct measures of their actual learning gains suggest the two forms of measurement are only weakly related (Bowman, 2010; 2011; Porter, 2012). Combined, these findings suggest that using SRLG as the main indication of student learning in engagement-outcome research may produce misleading results.

This SRLG issue is particularly troublesome for engagement-outcome literature. If in fact, these estimates are inaccurate, the current literature on student engagement may not accurately describe the relationship between engagement and student learning.

Pascarella et al. (2010), acknowledged the possibility of this issue stating, “We have, at present, very little internally valid evidence with respect to the predictive validity of the NSSE (p. 7).” They further voice concern that these findings are problematic if
institutions are expected to use NSSE data to make decisions to enhance student learning (Pascarella et al., 2010).

This is not to say that SRLG should be avoided entirely. While some uses may be inappropriate, there are many instances where self-reported measures have the advantage over direct measures. First, self-reports have the capability to measure constructs which may not be observable, something difficult to do with direct measures. Second, research has shown that while less precise than DMLG, SRLG may still provide a general indication of student learning (Dumont & Troelstrup, 1980; Gonyea, 2005; Kuh, 2001; Pace, 1984; Pike, 1995; 1996; Sitzman et al., 2010; Zilvinskis et al., 2015).

If this is true, SRLG may be sufficient for practical use where practitioners only need to detect general trends in learning and engagement. Additionally, if self-reported gains are somewhat representative of actual learning, it is unlikely that engagement-outcome research findings are entirely incorrect. More likely, if there are differences between SRLG and DMLG it would not impact the direction of their relationship with engagement but rather the strength of those relationships.

No doubt there are continued concerns regarding the issue of SRLG’s validity of their scores. The question is not whether SRLG’s are highly accurate – they are not – but rather to what degree do they represent the construct they are intended to measure. I suggest that more research comparing SRLG to DMLG is beneficial in engagement research for two reasons. First, comparing SRLG and DMLG may provide some indication of the degree of accuracy of SRLG and thus the accuracy of current research on engagement and learning outcomes. Second, if research using DMLG and SRLG produce comparable results, we may be confident that SRLG may roughly represent
student learning. If this is the case, researchers and practitioners could more confidently use the more convenient SRLG as a proxy for DMLG.

**Alignment of Measures.** Pike (1995) noted that the strength of relationships between self-reported measures and direct measures may be dependent on whether the content they measure is similar. While these two methods may claim to measure the same construct, differences in how broad the question are or what parts of the construct they measure may influence their relationships to other constructs (Pike, 1995). Pace (1985) noted that self-reports often ask broad and vague questions meant to measure a wide range of the construct. Conversely, direct measures tend to use specific items but are often limited in how much of the construct they can measure. According to Pike (1995), such differences in content attenuate relationships between the two despite measuring the same construct.

In a very similar way, the strength of relationships between engagement and learning may be influenced by the alignment of constructs in the measures. While engagement and outcomes are different constructs, their relationship is dependent upon whether (1) their content is related and (2) whether they share the same degree of specificity or generality.

Regarding related constructs, theoretically, there are a limited set of engagement practices that may contribute to SLOs. In other words, not every activity a student engages in will necessarily contribute to their learning in every area. If the content of the engagement subscale does not include practices that influence the outcome, there is no reason to expect that they would be related. Also, if an engagement subscale measures a mix of behaviors - some related to outcomes and some that are not, the relationship will
not be as strong as it would have only including behaviors related to it.

Additionally, the relationships between engagement and outcomes may be affected by how aligned they are in their specificity and content. Measures of engagement and outcomes may range from detailed, specific behaviors, to broad, sweeping generalizations about behavior or learning. Engagement subscales may broadly ask about student behavior regarding their engagement or may focus on specific aspects of quantitative reasoning. Likewise, outcomes measures may attempt to measure students’ general understanding of algebra or something as specific as their ability to read a graph.

Differences in specificity alignment may manifest in engagement-outcome relationships in many ways: broad engagement to broad gains, broad engagement to specific gains, specific engagement to broad gains, and specific engagement to specific gains. Theoretically, the strongest relationships between engagement and outcomes should be when both are broad, or both are specific.

Broad measures of engagement (NSSE) and broad learning (typically SRLG) may correlate relatively strongly because both measure their constructs in the same way. However, as discussed, broad measures are less prescriptive when used as a tool for improvement. Alternatively, the use of broad engagement and specific outcomes (DMLG) may lead to weakened relationships due to differences in the specificity with which they measure their constructs.

Finally, if there is a relationship between engagement and learning we might expect that specific engagement and specific outcomes would be the strongest. The shift from the original NSSE to the updated version provides an interesting example of the
impact a change in specificity may have on relationships. The updates to the NSSE involved an increase in the specificity of engagement subscales through the creation of the NSSE indicators. Zilvinskas, Masseria, and Pike (2017) conducted a study comparing the original NSSE themes* to the more specific NSSE 2.0 indicators. Their results suggested that the NSSE 2.0 provided greater insight into engagement-outcome relationships (Zilvinskis et al., 2017). Other studies breaking the NSSE into specific subscales have also found stronger and more practically applicable relationships (Pike, 2006a; 2006b; Zilvinskis et al., 2017). Based on these findings, it seems likely that using a more specific measure of engagement would lead to stronger engagement-learning relationships.

**Summary**

Student engagement has become a common multi-dimensional construct used to evaluate quality in higher education. Its popularity is driven by its purported relationship with student learning. The NSSE is the most prevalent of these measures, being used for institutional improvement, accountability, and research purposes. The validity of NSSE’s data for these uses has been mixed. The NSSE has been used to explore the engagement-outcome relationships but has predominantly relied upon SRLG. Such use of self-reported gains has been questioned, as it may produce misleading engagement-outcome results. While it is unlikely that the SRLG research has described a wholly inaccurate representation of the relationship between engagement and student learning, the precision of this evidence is dubious. The accuracy of the research’s representation of these relationships may be affected by two measurement issues. First, results may be influenced by whether learning gains are measured directly or collected via self-report.
Second, the alignment of the content measured by each instrument or subscale may influence the strength of the relationship. Thus, I argue that the utility of the information and relationships we observe from measures of engagement and learning gains depends on how the two are measured.

Taking these measurement issues into consideration, I propose that engagement-outcome research may benefit from an examination of the relationships between SRLG and DMLG. Further, an exploration of the alignment in content between SRLG and DMLG outcomes and measures of engagement may provide some insight concerning the strength of engagement-outcomes relationships. Finally, further light may be shed on engagement-outcome relationships through an exploration of the content alignment of engagement and outcomes measures.

In this study, it is my intent to explore how the methods used to measure engagement and student learning impact the observed relationship between the two meta-constructs. In addition, I intend to provide an example of how institutions and researchers may use direct measures of student learning in engagement research. The research questions I will address in this study include:

- **RQ1**: How does students’ general self-reported learning differ from specific self-reported learning?
- **RQ2**: Do students’ direct measures of learning gains differ depending on their self-reported learning gains?
- **RQ3**: To what degree do measures of general engagement and quantitative engagement differ from each other?
• RQ4: What are the relationships between engagement and quantitative reasoning learning gains?

• RQ5: Does general engagement predict quantitative reasoning learning gains as well as specific engagement?

• RQ6: Do general and specific engagement predict learning differently for each measure of learning?

• RQ7: Does the relationship between the most commonly used measures of learning and engagement (general SRLG and general engagement) differ from the relationship using the ideal forms of the engagement and learning measures (DMLG and specific engagement)?
CHAPTER THREE

Methodology

In this study, I explored how methods used to measure engagement and student learning affect the relationships observed between engagement and student learning gains. More specifically, this study emphasized how differing measurement techniques influenced the observed relationships between general and quantitative reasoning specific engagement and students’ learning of quantitative reasoning. In the following sections, I describe the participants, instruments, and procedures.

Participants

Study participants were a selection of undergraduate students at James Madison University (JMU). At JMU, before the beginning of their first semester in August, all first-year students are required to participate in a university wide assessment day. This will be referred to as the “Time 1” testing. Students are randomly assigned to rooms where a selection of tests and assessments are administered. Students who have acquired 45 to 70 credit hours are also required to attend a university wide assessment day in February. This will be referred to as the “Time 2” testing. Participants in this study were those who completed the Natural World-9 assessment during assessment day as freshman in August 2015, 2016, or 2017 and again as students with 45-70 credit hours in February 2018.

At February 2018 assessment day, the average age of students in the final dataset was 19.92 ($SD =0.39$). Consistent with JMU’s demographics, a greater proportion of participants were female than male. Participants predominately identified as Caucasian (84.7%). Additional participant demographics can be found in Table 4.
Measures

In measuring student learning and engagement, I emphasize the influence of two dimensions of measurement: specificity and directness. With respect to specificity, engagement was measured generally and then measured using items worded to measure engagement specific to quantitative reasoning. General engagement was operationalized using several subscales from the National Survey of Student Engagement. Specific measures of student engagement were developed by adapting the same subscales from the NSSE to emphasize quantitative reasoning (Table 3).

With regards to directness, student learning was measured in three ways. General self-reported learning gains (general SRLG), specific self-reported learning gains (Specific SRLG), and direct measures of quantitative reasoning learning gains (DMLG). General SRLG of quantitative reasoning was measured using an item borrowed from the NSSE. The measure of specific self-reported learning in quantitative reasoning was adapted from James Madison University’s general education objectives. The direct measure of quantitative reasoning learning gains (DMLG) was calculated by subtracting time-one scores from time-two scores on a measure of quantitative reasoning. To help ensure that our data were representative of students’ actual beliefs and abilities, a measure of test taking effort was used to filter students who put little effort into completing the DMLG.

**NSSE: Perceived Quantitative Learning Gains (General Self-reported Learning).** NSSE perceived gain items are used to indicate to educators how frequently students engage in effective educational practices. The NSSE asks students to estimate how much their institution has contributed to learning and developing in seven areas: writing, speaking, critical thinking, analyzing statistical and numerical information,
work-related knowledge/skills, personal values, ethics, and working with others.

Students respond to these questions on a four-point scale: very much, quite a bit, some, very little. NSSE recommends converting these responses to a 60-point scale, with “Very little” receiving a score of zero and increasing by increments of 20 for each increase along the four-point scale (i.e., Very little=0, Some=20, Quite a bit =40, Very much =60). One item from the NSSE was used as a measure of general QR SRLG. Because this item is like those used in other studies to measure self-reported gains, this study uses the question as a general measure of students’ SRLG in quantitative reasoning (Table 5; Carini et al., 2006; Pascarella, et al., 2009; 2010). This measure will be referred to as NSSE: Self-reported Learning Gains (NSSE: SRLG).

Quantitative Reasoning Self-Reported Learning Gain Scale (Specific Self-Reported Learning). One potential challenge of comparing SRLG to direct measures is an inconsistency in the range of content the instrument is designed to measure. To address this issue, I created a self-report learning instrument designed to align with James Madison University’s quantitative reasoning objectives.

James Madison University’s “Natural World” general education program has two types of objectives, those dealing with scientific reasoning and those emphasizing quantitative reasoning. As the direct measure used in this study was designed to assess whether the Natural World quantitative reasoning objectives were met, the two objectives were adapted to be a self-reported measure of learning gains on quantitative reasoning (see Table 6). Note that the wording of the instrument’s items is closely aligned with JMU’s quantitative reasoning objectives. Because the direct measure was created specifically to measure the objectives (Sundre, 2008), theoretically the DMLG and the
specific quantitative reasoning SRLG should be closely aligned. The instrument composed of the adapted objectives will be referred to as the Quantitative Reasoning Self-Reported Learning Gain Scale (QR-SRLG).

As with the NSSE, the QR-SRLG scale asks students to report how much their experience at the university has contributed to their knowledge, skills, and development on the objectives. The QR-SRLG’s response scale was intentionally designed to be the same as the NSSE’s: a four-point scale with the options, very much, quite a bit, some, very little. Students’ scores were converted to a 60-point scale, with “Very Little” receiving a score of 0 and increasing by increments of 20 for each increase along the four-point scale (i.e., Very little=0, Some=20, Quite a bit =40, Very much =60).

Quantitative Reasoning-9 (Direct Measure). The Quantitative Reasoning 9 (QR-9) is one of two subscales administered as part of the assessment for James Madison University’s “Natural World” general education program. The Natural World 9 (NW-9) was designed at JMU to directly measure student learning outcomes relevant to general education objectives for scientific investigation of the Natural World (Hathcoat, Sundre, & Johnston, 2015; Sundre, 2008; Sundre & Thelk, 2010). The NW-9 contains 66 questions divided between QR-9, and the Scientific Reasoning-9 (SR-9). In this study I only used data from the QR-9.

Since 2001 the NW-9 has undergone extensive study and development to enable sound psychometric properties, content alignment, and ease of use for students (Hathcoat, et al., 2015; Sundre, 2008; Sundre & Thelk, 2010). The NW-9 has been successfully used at other four-year institutions and has been found to be related to students’ scores on tests such as the ACT (Sundre & Thelk, 2010). Additionally, research has demonstrated
the NW-9’s ability to measure changes in quantitative ability over time (Hathcoat et al., 2015).

The QR-9 portion of the NW-9 consists of 26 multiple-choice items that are meant to measure students’ ability to critically think about quantitative information (Table 7; Sundre, 2008). Items are designed to have one or two correct answers and are scored as correct or incorrect. Students’ final scores are calculated as the number of items they correctly respond to. Historically, the reliability of data collected using the QR-9 subscale has been between .60 and .70 (Sundre & Thelk, 2010).

Engagement Measures. Two measures of engagement were used in this study, one measuring general engagement and the other measuring engagement specific to quantitative reasoning. Both measures were adapted from the NSSE1. Recall that the NSSE is intended to measure “the time and effort students devote to activities that are empirically linked to desired outcomes of college and what institutions do to induce students to participate in these activities” (Kuh, 2001; 2009) The NSSE divides student engagement into ten subscales (Table 8; NSSE: NSSE: Engagement indicators, 2017). Seven subscales were selected to be used in both measures of student engagement. These seven subscales were selected due to the ease with which they could be adapted for quantitative reasoning specificity. The subscales excluded from the engagement measures were those that contained items that were not easily or meaningfully adapted to be specific to quantitative reasoning (e.g., People from a race or ethnicity other than your own; People from an economic background other than your own).

In the general engagement measure, the wording of the seven selected NSSE

---

1 Permission for the use and adaptation of NSSE survey items was obtained from the National Survey of Student Engagement Indiana University
subscales was unmodified. In the specific measure of engagement, the subscales were altered for the context of quantitative reasoning by adding quantitative wording to each item (See Table 3 for a comparison of general and specific engagement measures). Note that to ensure that general engagement did not specifically measure quantitative reasoning, the NSSE quantitative reasoning subscale was excluded from the analyses of the engagement measures.

**NSSE: Engagement (General Engagement).** The general measure of student engagement was comprised of seven subscales totaling 25 items. No modifications were made to the wording of the items. As a result, this instrument provided an indication of students’ engagement not specific to quantitative reasoning. Participant responses were limited to a four-point scale, the wording of which varied depending upon whether the question asked about the students’ behavior or the school’s environment (See Table 9; Dumford & Rocconi, 2015). Students’ engagement scores were converted from the point scale to values of 0, 20,40, and 60, with the lowest level of engagement receiving a score of zero and increasing by increments of 20 (e.g., Never=0, Sometimes=20, Often =40, Very often =60; NSSE: Engagement indicators, 2017). Students’ engagement scores were then averaged within their respective subscales to provide seven general engagement scores. These scores were then averaged to form an overall engagement score.

**Adapted NSSE: Quantitative Reasoning Engagement (Specific Engagement).**

To measure student engagement specific to quantitative reasoning, I adapted the subscales used to measure general engagement, adding language relevant to quantitative reasoning. However, this instrument is distinct from the general engagement measure in
that it asks about behaviors that are more directly aligned with JMU’s quantitative general education program objectives. The adapted measure of engagement contained 25 items across seven subscales. To maintain similarity between measures of general and specific engagement no words were removed from the original NSSE items (Table 10; Appendix A). Rather, words were added to each question to make them relevant to quantitative reasoning. Each question retained its response format, giving students four response options. Students’ engagement scores were converted to scores of 0, 20, 40 or 60, with the lowest level of engagement receiving a score of zero and increasing by increments of 20 (i.e., Never=0, Sometimes=20, Often =40, Very often =60). Student’s engagement scores were then averaged within their subscales.

**Student Opinion Survey.** The student opinion scale (SOS) is an instrument designed to assess students’ test-taking motivation (Sundre, 2008). The SOS consists of two subscales one measuring effort students put into the test and other measuring the importance students place on the test. Each subscale contains five Likert-type items ranging from one (strongly disagree) to five (strongly agree). The range of possible scores on each subscale ranges from five to twenty-five (Thelk, Sundre, Horst, & Finney, 2009).

In this study, the effort subscale of the SOS was used to determine which students had put enough effort into the tests for scores to accurately represent their ability (Table 11; Thelk et al., 2009; Sundre, Thelk, & Wigtil, 2008; Sundre & Thelk, 2010). Data collected using the SOS has been shown to have good validity and strong reliability (Sundre, 2008; Sundre & Moore, 2002; Thelk, et al., 2009). In this study, the SOS was administered following the NW-9. Research has found that student motivation measured
by the SOS can be used to predict students’ performance on the NW-9 (Finney et al., 2016). This finding suggests that students who are less motivated perform poorly on the NW-9, and thus their data may not be an accurate representation of their ability (Finney et al., 2016). Because of this concern, the SOS will be used to filter out students with low motivation (a score less than 13) during the pre-test or Time 2 administrations of the direct measure (QR-9).

**Procedures**

In this study all participants completed the same measure of quantitative reasoning (QR-9) and motivation (SOS) at two time points. Participants completed Time 1 before the beginning of their first semester at JMU. Thus, data for Time 1 was collected in fall 2015, fall 2016, and fall 2017. Data for Time 2 was collected during spring 2018. During Time 2 participants completed five measures. First, participants were given one of two measures of engagement, the KES-A or KES-B, depending on which testing room they were assigned to. Thus, the dataset contained 178 participants who completed the general measure of engagement (KES-A) and 125 participants who competed the QR specific measure of engagement (KES-B). At the end of each measure of engagement all participants completed identical measures of general QR SRLG followed by a measure of specific QR SRLG. All participants were then given the NW-9 (which contains the QR-9), followed by the SOS.

**Missing Data.** Gathering direct learning gains scores on quantitative reasoning required that participants complete the QR-9 at two time points: in fall 2015 and spring 2018, fall 2016 and spring 2018 or fall 2017 and spring 2018. Recall that Time 2 is conducted when students have acquired 45-70 credits. Thus, some students matched this
criterion in the spring of their 1st, 2nd, or 3rd years. Of the 585 participants who participated in the study, 413 had data on the KES-A or KES-B and at test Time 1 and test Time 2. Participants who did not have complete QR-9 data or SOS data at either time point were excluded from the analysis. Additionally, those missing data on the KES-A or KES-B were removed from the dataset. The removal of missing data resulted in a sample size of 360 participants.

**Motivation Filtering.** Recall that there were no consequences for students’ performance on the direct measure of quantitative reasoning (QR-9). Therefore, students may not have been motivated to perform their best on the QR-9. Specifically, students who gave little effort likely produced scores that reflected low effort as opposed to their true quantitative reasoning proficiency. For the sake of validity, students with an SOS motivation score below 13 (out of 25) during either or both administrations of the NW-9 were excluded (Sundre & Wise, 2003; Swerdzewski, Harmes, & Finney, 2011; Wise & DeMars, 2005; Wise & Kong, 2005). Researchers have suggested that when students have low test motivation their data may not be reasonably assumed as representative of their true ability (Thelk et al., 2009; Wise & DeMars, 2005; Wise & Kong, 2005). The dataset prior to motivation filtering included 360 students. Mean student effort scores were 19.41 (SD= 3.37) at Time 1 and 18.74 (SD=3.38) at Time 2 (the possible range for effort scores is 5 to 25). After removing students with missing data and effort scores, 313 participants remained. The participants in the final dataset had a mean motivation score of 19.94 (SD=2.90) at Time 1 and 19.33 (SD =2.85) at Time 2.

**Analysis**
In this study I compared relationships among measures of engagement and learning. However, in doing so there was a methodological conundrum. While all constructs in this study are theoretically continuous, the self-reported measures of learning (SRLG) and engagement use a four-point scale, which produces non-continuous data. For example, the NSSE self-reported learning gains measure ranges from a score of one to four (or 0 to 60 when converted). Despite these data conditions, most previous NSSE research has treated such variables as continuous (Carini et al., 2006; Gonyea, 2006; Pascarella et al. 2009; 2010; Kuh, Pace & Vesper, 1997).

The conundrum lies in whether to treat the four-point scales as continuous – and be more consistent with past research – or treat the data as categorical, which would be more methodologically sound. I chose a hybrid approach.

For the first two research questions (RQ1 and RQ2), treating the four-point scales as categorical was straightforward in the analyses, and thus I used categorical approaches. For RQ3 through RQ7, the analyses would have been more complicated using categorical analyses. Thus, I chose the traditional, albeit flawed, approach of treating these items as continuous. I revisit the implications of these methodological decisions in the discussion.

**RQ1: How does students’ general self-reported learning differ from specific self-reported learning?** This research question addresses whether student responses differed across the two measures of SRLG. I compared the median scores of each instrument using a non-parametric test, the Wilcoxon Signed Ranked Test. This test can be thought of as a non-parametric counterpart to a paired-samples t-test. This test determines if students’ self-report of quantitative reasoning depends on whether the items are general with respect to quantitative reasoning or specific related to JMU’s
quantitative reasoning student learning outcomes.

**RQ2: Do students’ direct measures of learning gains differ depending on their self-reported learning gains?** This research question is intended to determine whether DMLG differed depending on student’s responses to measures of SRLG. This research question was addressed by comparing direct measures of student gains across different levels of SRLG. If self-reported learning aligns with actual learning gains on the QR-9 I would expect that those who indicated learning little would have the lowest gain scores while those who indicated learning a great deal should have the greatest gain scores. These groups were compared using two, one-way, independent-samples ANOVAs. The levels of the independent variable being the response categories of the self-reported measures. The dependent variable was students’ change scores on the QR-9. By framing the analyses in this way, we can determine if students who self-reported on each scale differed from one another in their gains as indicated by the QR-9. In addition to determining whether change scores differ depending on self-reported learning, I also observed the effect sizes of group differences. This analysis helped determine the magnitude of direct measure gain score to particular levels of self-reported learning gain (e.g., little, some)

**RQ3: To what degree do measures of general engagement and quantitative engagement differ from each other?** To address this question, I compared the engagement scores of students who completed the general engagement measure to students who completed the quantitative reasoning specific measure of engagement. This comparison was done using a simple independent t-test. Note that scores on engagement subscales reflected several items as opposed to the SRLG.
RQ4: What are the relationships between engagement and quantitative reasoning learning gains? Simple correlations were calculated to explore the relationship between each subscale of each measure of engagement and the different measures of learning. These analyses indicate the strength of the linear relationship for each combination of measures. That is, these analyses describe the strength of the relationships between each engagement measure and each learning measure.

RQ5: Does general engagement predict quantitative reasoning learning gains as well as specific engagement? To address this question, I compared the two measures of engagement on their relationship to each measure of learning. For example, I compared the relationship between general engagement and NSSE: SRLG to the relationship between specific engagement and NSSE: SRLG. To test for statistical significance, I used Fisher’s r to z transformation followed by Steiger’s test of independent correlations. These analyses help determine if the two measures of engagement predict learning differently. In total there were three comparisons.

- General Engagement-General SRLG to Specific Engagement-General SRLG
- General Engagement-Specific SRLG to Specific Engagement-Specific SRLG
- General Engagement-DMLG to Specific Engagement-DMLG

RQ6: Does the strength of relationship between engagement and learning differ depending on the type of learning measure? To address this question, I compared various pair-wise relationships of engagement and learning. For example, I compared the relationship between general engagement and NSSE: SRLG to the relationship between general engagement and QR-9: SRLG. To test for statistical significance I used Fisher’s r to z transformation followed by Steiger’s test of dependent
correlations. These analyses helped determine if the relationship between engagement and learning is dependent on the measure of learning used. There will be six comparisons in total:

- General SRLG-General Engagement to Specific SRLG-General Engagement
- General SRLG-General Engagement to DMLG-General Engagement
- Specific SRLG-General Engagement to DMLG-General Engagement
- General SRLG-Specific Engagement to Specific SRLG-Specific Engagement
- General SRLG-Specific Engagement to DMLG-Specific Engagement
- Specific SRLG-Specific Engagement to DMLG-Specific Engagement

These comparisons involved calculating the correlations between each measure of engagement and each measure of learning. Because general and specific engagement measures were given to separate samples, comparison of engagement-learning relationships were conducted for both measures of engagement. Thus, for general engagement the correlation of general self-reported learning gains with engagement was compared to that of specific self-reported learning gains as well as direct measured learning gains. Then, the correlation of specific self-reported learning gains and general engagement will be compared to that of direct measured learning gains. The same three comparisons of learning measures were then conducted with the measure of specific engagement.

RQ7: Does the relationship between the most commonly used measures of learning and engagement (general SRLG and general engagement) differ from the relationship using the ideal forms of the engagement and learning measures (DMLG and specific engagement)? To address this question, compared the relationship
between general engagement and NSSE: SRLG to the relationship between quantitative specific engagement and QR-9: DMLG. This was done by converting the correlations using Fisher’s r to z transformation followed by Steiger’s test of independent correlations. These analyses tell us whether general engagement is more strongly related to self-reported learning than quantitative specific reasoning is to learning gains on a direct measure of quantitative reasoning.
CHAPTER FOUR

Results

RQ1: General vs. Specific Self-reported Learning Gains

Descriptives. Transformed general self-reports of quantitative reasoning learning gains were measured using a single question with responses ranging from zero to 60 with zero being equivalent to a response of “very little”, 20 being equivalent to “some”, 40 being “quite a bit” and 60 being “very much”. After cleaning and filtering, the average general self-reported learning gain was 22.68 ($SD=17.91$), which is closest to a response of “some”. Reliability was not calculated, as this was a single item instrument.

Specific self-reported learning gains were measured with two items. On each item students responded on a scale of one to four. In order to put the specific SRLG on the same scale as the general SRLG, students’ responses on the specific SRLG were used to assign students to groups corresponding to the four-point scale. Students whose total score ranged from two to three were assigned a score of zero (learned very little) those with a total score of four were assigned a score of 20 (learned some). Those whose total score was between five and six were assigned a score of 40 (learned quite a bit) and those between seven and eight were assigned a score of 60 (learned very much). This transformation resulted in specific self-reported learning gains being on the same 0, 20, 40, 60 scale as general SRLG. The average of the rescaled specific SRLG was 25.68 ($SD=18.93$), slightly higher than the general SRLG. The reliability of the specific self-reported learning gains was moderate at $\alpha=.75$.

Analysis. A Wilcoxon Signed Ranked Test was conducted to compare general self-reported learning gains to the specific self-reported learning gains. The median of
both measures of SRLG was 20. The test revealed a statistically significant difference between the two measures of self-reported learning (Z (183, 130)=-3.303, p=.001).

These findings indicate that students generally reported greater learning on the specific measure of self-reported learning than they did on the general measure of self-reported learning.

**Research Question Two: Direct-measured Learning Gains vs. General and Specific Self-reported Learning Gains**

**Descriptives.** To calculate the direct measure of student quantitative reasoning learning gains, students’ Time 1 scores on the QR-9 were subtracted from their Time 2 scores. QR-9 change scores could range from –26 to +26 with positive scores indicating improved performance from Time 1 to Time 2. After cleaning and filtering, the average Time 1 QR-9 score was 17.65 (SD=3.34). The reliability of student scores on the QR-9 at Time 1 was α=.57. At Time 2, the average QR-9 score was 18.59 (SD=3.62). The reliability of the QR-9 at Time 2 was higher but still modest at α=.66. The change in student performance on the QR-9 ranged from one to twenty-eight from Time 1 to Time 2 with a mean of .936 (SD=3.63).

**Analysis.** A one-way between-subjects ANOVA was computed to determine whether students’ general self-reports of their learning corresponded to differences in the direct measures of their learning gains. The QR-9 difference scores were treated as the dependent variable and students’ general self-reported learning was used as the grouping variable with four levels: learned very little, learned some, learned quite a bit, learned very much. The results showed no statistically significant difference of QR-9 learning gains based on students’ self-reported learning gains (F(3,309)= 2.204, p=.08,
Another one-way between-subjects ANOVA was computed to determine whether students’ specific self-reports of their learning corresponded to differences in the direct measures of their learning gains. The QR-9 difference scores were treated as the dependent variable and students’ specific rescaled self-reported learning was used as the grouping variable. The results showed no statistically significant difference of QR-9 learning gains based on students’ self-reported learning gains ($F(3,309)=0.935$, $p=.424$, $\eta^2_{par}=.009$; Table 13).

**Research Question Three: General Engagement vs. Specific Engagement**

**Descriptives.** General engagement was measured using a set of unmodified subscales from the NSSE. This instrument contained 25 questions on a scale of one to four. As discussed earlier these scores were transformed to be on a scale of zero to 60, with one corresponding to zero, two to 20, three to 40, and four to 60. Participants’ summed total engagement score could range from zero to 1,500 (i.e., $25 \times 60$). The average total general engagement score was 548.42 ($SD=210.88$). The reliability of the general engagement scale was high at $\alpha=.889$. Specific engagement was measured using the same subscales and general engagement. However, each question was slightly modified to include wording specific to quantitative reasoning. Thus, the quantitative reasoning specific measure of engagement was also comprised of 25 questions and total scores could range from zero to 1,500. The average specific engagement score was 632.15 ($SD=271.01$). The reliability of the general engagement scale was high, at $\alpha=.929$.

**Analysis.** A one-way between subjects $t$-test was conducted to determine if
general engagement differed from quantitative specific engagement scores. The test revealed a statistically significant difference between the two measures of engagement ($t(311)=3.072$, $p=.002$, $d=0.344$). Quantitative specific engagement scores were, on average, 83.73 points higher (on a 1,500 point scale) than general engagement scores.

**Research Question Four: Relationships between Engagement and Learning**

**Analysis.** This research question was addressed through the calculation of simple Pearson correlations among the key variables used in this study. The correlations and their $p$-values can be seen in Table 14 and Table 15. Generally, the self-reported measures of learning (general and specific) were found to positively, significantly, and moderately relate to each other and both measures of engagement. The direct measure of student learning was found to have no significant relationship with specific self-reported learning or engagement. However, the direct measure of student learning (QR-9 difference score) was found to be significantly, negatively, and weakly related to general self-reported learning. In other words, students who self-report higher levels of general engagement were more likely to have learning decreases (i.e., negative learning gains). Note, however, that the effect size for the negative correlation was small ($R^2=.02$). In other words, only 2% of the variance in learning gains could be explained by self-reported general engagement (and vice versa).

**Research Question Five: General Engagement vs. Specific Engagement in Predicting Learning**

**Analysis.** To address this question, correlations were transformed to $z$-scores using Fisher’s $r$ to $z$ transformation. Then, the transformed correlations were compared using a simple $z$-test to determine whether the measures of engagement predict
quantitative reasoning learning gains differently (Cohen, Cohen, West & Akin, 2003). Three comparisons were conducted. To control for inflated Type I error a Bonferroni adjustment was used to ensure $\alpha=.05$ for the set of comparisons. Thus, the cut-off for statistical significance was set at $\alpha=.016$ (i.e., $0.05/3$).

Preacher’s (2002) tool for the comparison of independent correlations follows the procedure described above and was used to compare the relationship of general engagement ($r(181)=.376$, $p<.001$) and quantitative reasoning specific engagement ($r(128)=.489$, $p<.001$) to general self-reported learning. The test indicated no statistically significant difference between the two relationships ($z(183, 130)=-1.202$, $p=.229$).

Another test was conducted comparing general ($r(181)=.520$, $p<.001$) and specific ($r(128)=.586$, $p<.001$) engagement on their relationships to specific self-reported learning. The $z$-test of independent correlations indicated no statistically significant difference between general and specific engagement in their relationship to specific self-reported learning gains ($z(183, 130)=-0.822$, $p=.411$). Last, the relationships of general ($r(181)= -.036$, $p=.631$) and specific ($r(128)= .009$, $p=.921$) measures of engagement to the QR-9 change scores (direct measure) were compared using the $z$-test of independent correlations. Results showed no significant difference between general and quantitative engagement in predicting QR-9 change scores ($z(183, 130)=0.388$, $p=.698$).

**Research Question Six: Differences Between Measures of Learning in Their Relation to Engagement**

**Analysis.** To address this question Steiger’s test of dependent correlations for correlations sharing one variable was used to compare each measure of learning on their relationship with engagement (Steiger, 1980). This analysis was done using software
made available by Lee and Preacher (2013). Because the two measures of engagement were not equivalent, measures of learning were compared separately for each measure of engagement. Prior to comparison, correlations were standardized using Fisher’s $r$ to $z$ transformation as described previously. Following transformation, a total of six comparisons were conducted. The first three comparisons were conducted using the general measure of engagement the last three comparisons were conducted using the quantitative reasoning specific measure of engagement. To control for inflated Type I error a Bonferroni adjustment was used to ensure $\alpha = .05$ for the six comparisons. Thus, the cut-off for statistical significance was set at $\alpha = .008$.

The first comparison tested for differences between general self-reported learning and specific self-reported learning in their relationship to general engagement. Steiger’s test indicated no significant differences in these relationships ($z(182) = -2.600, p = .009$). This tells us that general engagement was not more strongly related to specific self-reported learning ($r(181) = .520, p < .001$) than with general self-reported learning ($r(181) = .376, p < .001$). The second comparison tested for differences between general self-reported learning ($r(181) = .376, p < .001$) and QR-9 change scores ($r(181) = -.036, p = .631$) in their relationship to general engagement. Steiger’s test indicated a significant difference in these relationships ($z(182) = 3.834, p < .001$). This tells us that general engagement was more strongly related to general self-reported learning than the direct measure of learning gains. The third comparison tested for differences between specific self-reported learning ($r(181) = .520, p < .001$) and QR-9 change scores ($r(181) = -.036, p = .631$) in their relationship to general engagement. Steiger’s test showed significant differences in these relationships ($z(182) = 5.583, p < .001$). These results mean that
general engagement was more strongly related to specific self-reported learning than the direct measure of learning gains.

The fourth comparison tested for differences between the relationships of general self-reported learning \( (r(128)=.489, p<.001) \) and specific self-reported learning \( (r(128)=.586, p<.001) \) to quantitative reasoning specific engagement. Stieger’s test indicated no significant differences in these relationships \( (z(129)=-0.868, p=.385) \). The fifth comparison tested for differences between general self-reported learning \( (r(128)=.489, p<.001) \) and QR-9 change scores \( (r(128)=.009, p<.921) \) in their relationship to quantitative reasoning specific engagement. Steiger’s test indicated a significant difference in these relationships \( (z(129)=4.033, p<.001) \). These results indicate that general engagement was more strongly related to general self-reported learning than the direct measure of learning gains. The last comparison tested for differences in the relationship between specific self-reported learning and QR specific engagement \( (r(128)=.586, p<.001) \) and the relationship between QR-9 change scores and QR specific engagement \( (r(128)=.009, p<.921) \). Steiger’s test indicated a significant difference in these relationships \( (z(129)=5.302, p<.001) \). In other words, the relationship between specific self-reported learning and QR specific engagement was stronger than the relationship between QR-9 change scores and QR specific engagement.

**Research Question Seven: Engagement-Learning Relationships Using Common vs. Ideal Methodology**

**Analysis.** To address this research question, the relationship between general engagement and general self-reported learning \( (r(181)=.376, p<.001) \) were compared to the relationship between quantitative reasoning specific engagement and direct-measured
student learning ($r(128)= .009, p=.921$). This was done by transforming each correlation using Fisher’s $r$ to $z$ transformation and conducting a single comparison using a $z$-test of independent correlations, as recommended by Cohen et al, (2003). The analysis was conducted using Preacher’s (2002) tool for the comparison of independent correlations. The analysis revealed a significant difference between the engagement-learning relationships of the ideal vs common measurement methods ($z (183, 130)= 3.334, p<.001$). These results indicate that the relationship between general engagement and general self-reported learning gains (common method) was larger than the relationship observed between QR specific engagement and the QR-9 change score (Ideal method).
CHAPTER FIVE

Discussion

Given engagement’s widespread use in higher education, it is prudent to understand the relationship between engagement and learning. While research has generally found support for small to moderate relationships between student engagement and learning (Christenson, Reschly, & Wylie 2012; Finn & Zimmer 2012; McCormick, 2008; Trowler, 2010), the methodologies used in these studies may be subject to question.

One concern is the use of self-reports to measure student learning gains (SRLG), the predominate method in engagement-learning research. In many circumstances students may not be able to objectively estimate how much they know or have learned (Porter 2011; Porter, Rumann, & Pontius, 2011). If student estimates of their learning are indeed inaccurate, then there is reason to question the accuracy of engagement-learning relationships revealed by research.

A second concern is the common practice of relating measures of general engagement to specific learning in subjects such as quantitative reasoning (i.e., a specificity mismatch). The problem behind this practice is that general engagement measures students’ engagement across a wide range of experiences. Relationships between general engagement and learning in a specific subject may be inaccurate because a portion of students’ engagement score is irrelevant to the learning subject. If this is the case, much of the research may actually be underestimating the relationships between engagement and learning.

The purpose of this study was to address each of these concerns and consider
alternative methods that may provide a more accurate representation of the relationship between engagement and learning. One goal of this study was to explore whether self-reported learning (SRLG) was related to direct measures of learning (DMLG). And relatedly, whether the relationships between engagement and learning were similar for self-reported learning and direct gain score measures of learning. A second goal of this study was to determine to what degree engagement-learning relationships differ when using a general measure of engagement and a learning subject-specific measure of engagement. Finally, this study compared engagement-learning relationships found using the most common methodology (general engagement and SRLG) to a proposed alternative (specific engagement and DMLG).

In the following sections I provide a summary of the findings for each research question as well as a discussion of what the findings may mean as a whole. Finally, I discuss implications of this study as well as provide suggestions for future research.

**Learning Question One: General Self-reported Learning and Specific Self-reported Learning**

Although self-report of student learning gains is treated as continuous for most of the research questions, for research question one I used categorical analyses to compare the two self-reported measures of learning. This was done because both measures involved were categorical. The findings of the analysis revealed a statistically significant difference between the typical student response to the general self-reported measure and the specific self-reported measure of quantitative reasoning learning gains. The general self-report was slightly lower than the specific self-report. As both self-reported measures ask about quantitative reasoning learning, these findings support the idea that
self-reports of learning may change depending on the degree of specificity with which they are measured.

**Research Question Two: Direct-measured Learning Gains vs. General and Specific Self-reported Learning Gains**

Research question two focused on determining whether students’ self-reported learning corresponded to their direct measured learning gains. Research has indicated that students may have difficulty estimating their learning gains. If this is the case, the use of indirect measures of student learning may be questionable in engagement-learning research. However, if self-reported learning and learning gains are strongly related it would suggest that self-reports may act as proxies for direct measured learning. Such findings would support the current literature and support the use of easy-to-administer self-reported measures.

The results for research question two indicated non-significant differences in direct measured learning gains (DMLG) between students who self-reported learning “very little”, “some”, “quite a bit”, and “a great deal”. That is, variance in DMLG could not be explained using students’ self-reported learning. In the context of this study the results suggest that students’ self-reports of learning do not consistently align with direct learning gains. These findings are in alignment with research showing students may not be able to accurately report their learning.

However, it is important to note that the QR-9, from which the direct measures of learning were calculated, demonstrated marginal reliability (~60). With such low reliability it is likely that any relationship that might be present would be attenuated.
Further complicating the reliability issue is that change scores of the direct measures were used, which can exacerbate unreliability.

Despite the less than optimal reliability, these findings suggest that the self-reported learning gains were not an accurate proxy for direct measured learning gains. To be a good proxy both measures should be strong to moderately related. Even with a reliability of .60 for single administrations and using change scores, one would expect to observe at least a small but discernible, statistically significant relationship if in reality the relationship was moderate; no such relationship was observed.

**Research Question Three: General Engagement vs. Specific Engagement**

Research question three explored whether the relationships between engagement and learning measures are dependent upon specificity alignment. Measures of engagement and outcomes may range from detailed, specific behaviors, to broad, sweeping generalizations about behavior or learning. Most often research has used broad and general measures of engagement while using specific measures of student learning. The analyses for research question three checked for differences between general engagement and quantitative reasoning specific engagement. Recall that each measure of engagement contained the same questions, drawn from the NSSE. However, the questions in the quantitative reasoning specific measure were slightly adapted to contain wording concerning quantitative reasoning. Each measure was administered to a different set of students. Results from the analysis indicated a significant difference between general engagement and specific engagement. More specifically, students’ average engagement was higher for the quantitative reasoning measure of engagement than the general measure of engagement.

Because quantitative reasoning engagement should be encapsulated within
general engagement, I might expect student responses to quantitative reasoning engagement to be lower than the general engagement scores. However, this was not the case. One possible explanation for this finding is that, because quantitative reasoning engagement is more specific it may be easier for students to recall specific instances of their engagement than when they are asked to consider their engagement across all subjects. It may also be possible that when students respond to questions about their engagement they do not do so by cumulating instances of their engagement but may instead be averaging their engagement across subjects. While either of these explanations may account for the difference between general and specific engagement, it is not possible to explain the pattern using data from this study.

Because random assignment was used, there was little reason to expect that the groups taking each test differed in their experiences at JMU. Thus, these results seem to suggest that the differences in engagement means may be a result of the specificity of the instrument and emphasis on quantitative reasoning.

**Research Question Four: Relationships between Engagement and Learning**

The analyses for research question four were intended to provide an overview of the correlations between each measure of engagement and each measure of learning. Results showed that both general and specific measures of engagement were significantly and positively related to the self-reported measures of learning. That is to say, students who self-reported higher engagement also self-reported greater learning gains. These findings are in line with much of the other research on engagement and student leaning. Much of the research on learning and engagement has found significant moderate relationships between engagement and learning measures. If self-reports of quantitative reasoning learning gains are good proxies of direct measures of quantitative reasoning
learning I might expect to see similar relationships between direct measured learning and engagement as I saw between self-reported learning and engagement. However, based on the findings of research question two, self-reported learning did not correspond to direct measures of quantitative reasoning learning gains. This finding was supported by a weak negative relationship between general self-report learning gains and direct measured learning gains. Thus, it is unsurprising that general engagement was found to be uncorrelated with direct measures of student learning. Additionally, quantitative reasoning specific engagement was also found to be unrelated to quantitative reasoning learning gains.

Combined with the findings from earlier research questions, these results suggest that the students’ self-reports of QR learning were not similar to students’ QR direct learning gains. Instead, it appears that the relationship between engagement and learning depended on the methods used. The use of methodology typical to engagement-learning research led to findings similar to those in other studies, a moderate positive significant relationship. However, when direct measures were used in place of self-reports, the relationship between learning and engagement disappeared. Note that this study’s finding was more severe than Pascarella et al 2009 and 2010, as they found (small?) but statistically significant relationship between direct measure gain scores and engagement.

**Research Question Five: General Engagement vs. Specific Engagement in Predicting Learning**

Based on the findings for research question one through four, the engagement-learning relationships appeared to differ depending on the methods and measures. Research question five asks whether general and QR specific measures of engagement
are differentially related to learning. Tests were conducted to compare the correlations between QR specific engagement and each measure of learning to the relationships between general engagement and each measure of learning. To conduct this test all correlations were transformed using Steiger’s r to z transformation and a Bonferroni’s adjustment was used to control for type one error. A total of three comparisons were conducted one for each measure of student learning (general self-report, specific self-report, and direct measure).

- General Engagement-General SRLG to Specific Engagement-General SRLG
- General Engagement-Specific SRLG to Specific Engagement-Specific SRLG
- General Engagement-DMLG to Specific Engagement-DMLG

Recall that some students completed the general measure of engagement while others took the specific measure. Because of this design, the correlations being compared came from two separate samples. Nevertheless, random assignment was used so we have no reason to believe that the two samples differed from one another systematically. Therefore, the significant difference between two measures of engagement seem to be due to the difference in their specificity rather than group differences in engagement.

While the two measures of engagement had significantly different means (research question three), the two measures of engagement did not differ in their relationships (strength and direction of correlation) to any measure of student learning. Quantitative reasoning-specific measures of engagement were not significantly different from the general measure of engagement. Thus, the expectation that quantitative reasoning measures of engagement should be more strongly related to measures of quantitative reasoning learning was not supported.
Research Question Six: Differences between Measures of Learning in their Relation to Engagement.

While research question five addressed differences between measures of engagement in their relation to student learning, research question six addressed differences between engagement-learning relationships across different measures of learning. Research question six was designed to test to what degree the relationship between engagement and learning is dependent on the measures used. Differences between engagement–learning relationships were tested for each measure of learning within each measure of engagement. This was done using six Steiger’s tests of dependent correlations listed in the bullets below.

- General SRLG-General Engagement to Specific SRLG-General Engagement
- General SRLG-General Engagement to DMLG-General Engagement
- Specific SRLG-General Engagement to DMLG-General Engagement
- General SRLG-Specific Engagement to Specific SRLG-Specific Engagement
- General SRLG-Specific Engagement to DMLG-Specific Engagement
- Specific SRLG-Specific Engagement to DMLG-Specific Engagement

The findings from the six comparisons showed that the general and specific self-reported measures [of learning] were not different in their relationship to either measure of engagement. This might be expected as both measures were self-reports intended to address the same construct. Although the direct measures of learning were intended to get at the same construct as the self-reports, research and findings in earlier research questions suggested a weak relation between self-reported learning and direct measured learning. This pattern was reflected in the comparison of each measure’s relation to
engagement. Regardless of the measure of engagement, both self-reported measures of engagement were significantly different than direct measures in the relationship to engagement. This is unsurprising, as earlier research questions revealed no significant relationship between direct measures of engagement but found significant relationships between engagement and self-reported learning measures. These findings seem to provide some justification for the argument that self-reports of student learning may not be equivalent to measures of their actual learning.

**Research Question Seven: Engagement-Learning Relationships Using Common vs. Ideal**

This research question addressed whether there were any discrepancies in engagement-learning relationships between the methodologies commonly used in the research versus what I proposed as an alternative, ideal technique. That is, this research question checked for differences in the engagement-learning relationship when using general engagement and a self-reported measure of QR learning versus using a QR specific measure of engagement and a direct measure of quantitative reasoning learning. Results indicated a significant difference in engagement-learning relationships between the two sets of measures. Thus, the measurements of learning used did significantly influence the relationship observed between engagement and learning. The difference observed may be unsurprising as measures of engagement were not found to differ from one another, but self-reported measures of learning were found to be significantly different from the direct measures of learning. Therefore, it seems probable that the differences between the engagement-learning relationships are largely a result of differences in the measures of learning used.
Implications

Overall, the results of this research suggest that the engagement-learning relationship is dependent on the measurement used. Although I expected that having a greater degree of specificity of engagement measures would lead to stronger relationships between engagement and learning, no differences were found in this study. Despite the lack of significant differences between engagement-learning relationships using different measures of engagement, it did generally appear that the relationship using quantitative reasoning specific measures of engagement were slightly elevated beyond the general measure of engagement. Thus, it may be worthwhile for future research to further explore how the specificity of the engagement measure may influence the observed relationships between engagement and learning.

While differing engagement measures did not appear to result in differences in engagement-learning relationships, the measure of learning did. If self-reported measures are accurate representations of student learning they should be strongly related to changes in students’ performance on a test of quantitative reasoning over time (i.e., direct measures of student learning gain). Additionally, if self-reports were good proxies of direct measures I would expect them to be similarly related to engagement. Generally, I found that the relationships using self-reported measures of learning were similar to each other. However, self-reported measures of learning were not strongly related to quantitative reasoning direct change scores over time. In fact, general self-reported learning had a weak negative correlation to gain scores. This suggests the possibility that those who reported greater learning were actually those who demonstrated the least gains in learning using direct measures. Additionally, self-reported measures of learning and
direct measures did not share a similar relationship with measures of student engagement. These findings seem to justify the concern that the use of self-reported measures of learning in engagement research may not necessarily provide the same, or even similar results as would be found using direct measures of learning gains.

If I assume that direct measures of student performance are representative of students’ actual ability, then my findings suggest that students’ self-reports of their QR learning may not be representative of their actual learning. Because these results are limited to a relatively narrow scope of quantitative reasoning learning and engagement they cannot be generalized to suggest that all self-reports of learning are inaccurate. However, these findings are in line with research which has cautioned against the use of self-reported learning gains.

More research is needed to understand how the use of direct and self-reported measures of learning influence the relationships observed between engagement and learning. The findings of this study demonstrated that direct measures and self-reported measures of quantitative reasoning learning are not necessarily related. This study, however, serves an example of the type of study needed to investigate such relationships.

The results of this study do not justify abandoning all use of self-reported measures of learning in engagement-learning research. Rather, these findings suggest practicing caution when using self-reported measures of student learning when studying the relationships between engagement and learning. Ideally, researchers should consider only using self-reported measures that have been found to correlate substantially with well validated direct measures of learning gains. In other words, research should ensure that student self-reports are reasonable proxies for more objective measures of learning
prior to using them. Finally, even if the research found self-reports to consistently lead to significantly different relationships than direct measures there may be ways to make self-reports more representative of direct measures. Techniques may be developed to help make students more accurate in their estimation, for example using anchors to have students recall their starting point and their current point to estimate learning over time (Finney, Putnam, & Boyd, 1998).

**Limitations and Future Directions**

While the findings from this study suggest that the methods used to measure engagement and learning may influence observed engagement-learning relationships, this is the only study I am aware of which has addressed this issue. Additionally, there are a number of methodological issues that warrant caution in the interpretation and generalization of this study’s findings.

I assumed my direct measure of quantitative reasoning (the QR-9) provided an accurate and psychometrically sound representation of students’ quantitative reasoning ability. However, the reliability of the QR-9 was marginal for both Time 1 and Time 2 measures. Such low reliability brings with it a number of issues including attenuation of relationships. Within this study gain scores computed form the QR-9 were related to other measures of learning as well as to measures of engagement. That being the case, it is likely that the observed relationships between the QR-9 and other measures were attenuated substantially due to the unreliability. With a major portion of this study involving the comparison of relationships the low reliability (and subsequent attenuation) may call into question the results of the comparisons. Thus, it is critical that researchers studying the influence of methods on engagement-learning relationships should be
careful to select measures with strong psychometric properties.

A second possible limitation was the use of simple gain scores. In this study, simple gain scores were calculated by subtracting Time 1 scores from Time 2 scores. Some have voiced concern with the use of simple gain scores, citing issues of floor and ceiling effects as well as problems with reliability (Cronbach & Furby, 1970; Overall & Woodward, 1975; Zumbo, 1999). Although scores in this study seemed to be well distributed, low reliability of the QR-9 makes it reasonable to conclude that the gain scores would also demonstrate low reliability. These issues may lead to the underestimation of participants’ actual gains. Thus, with the issues of low reliability, I cannot be confident that students’ gain scores on the QR-9 were representative of their actual learning over time. In the future, researchers may want to consider the use of more complex, but less problematic techniques for analyzing gain scores in engagement-learning research.

Another limitation of this study and of engagement research generally, is treatment of four-point scales as continuous. The National Survey of Student Engagement uses a four-point scale to measure engagement and student learning. As such, all measures, excluding the direct measure of student QR learning gains (QR-9), were on a four-point Likert scale. While the use of four-point scales are common in the engagement-learning research the appropriateness of their use as a continuous measure is questionable. Foremost, the four-point scales used by the NSSE often do not provide a full continuum of possible responses. That is, the scale for many items ranges from “Very Little” to “Very Much”. This provides students no opportunity to indicate a complete lack of engagement/growth or even negative engagement/growth. With a four-
point scale, there may also be an issue of restricted range, which might attenuate relationships between measures.

In the future, researchers should consider the possibility of treating NSSE engagement scales as ordinal, and thus using non-parametric analyses. They could also consider adding more points to the scale that might reflect zero and/or negative engagement and learning growth.

Finally, the self-reported measures of student learning were very brief, with the general measure comprised from one item (as used in the NSSE), and the specific measure comprised of two items. With a limited number of items in each of these scales, the findings in this study are limited to brief self-reports of learning. It would be interesting to explore whether relationships between self-reported meaning gains and direct measures of learning gains change across a larger range of specificity and number.

Conclusion

The assumption that engagement leads to student learning is fundamental in higher education. The purpose of this study was twofold. First, this study demonstrated the use of gain scores as measures of learning in engagement-learning research. Second, this study explored how methods used to measure engagement and quantitative reasoning learning gains may influence observed engagement-learning relationships. Specifically, I tested the influence of the specificity of the engagement measure as well as the specificity and response type on observed engagement-learning relationships. While specificity of engagement measures did not significantly influence the relationships, measures of learning did. Across measures of quantitative reasoning learning gains I found that self-reported learning measures were related to each other and were more strongly related to
engagement than direct measures of learning. These results suggest that self-reports may not provide the same results as direct measured learning. Assuming direct measures of gains represent a more objective view of students’ learning than self-report, these findings suggest that self-reports of quantitative reasoning may not accurately represent students’ actual learning. This research is valuable to the engagement literature as very little research has been conducted using change scores as measures of learning (Pascarella et al being the exception). This study provides an example of how researchers might further explore the engagement/learning relationship further.

Additionally, no research I am aware of has explored how measurement methodologies influence the observed-engagement learning relationships. Although there are many methodological concerns that limit the generalizability of my findings, I believe that this work provides a valuable framework by which the appropriateness of methods used in engagement-learning research may be assessed. It is my hope that researchers will consider and further explore the possible influence of instrument specificity and measurement method on engagement research.
Table 1
*Themes and indicators of the NSSE 2.0*

<table>
<thead>
<tr>
<th>Theme</th>
<th>Engagement Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of Academic Challenge</td>
<td>Higher-Order Learning</td>
</tr>
<tr>
<td></td>
<td>Reflective &amp; Integrative Learning</td>
</tr>
<tr>
<td></td>
<td>Learning Strategies</td>
</tr>
<tr>
<td></td>
<td>Quantitative Reasoning</td>
</tr>
<tr>
<td>Active and Collaborative Learning</td>
<td>Collaborative Learning</td>
</tr>
<tr>
<td></td>
<td>Discussions with Diverse Others</td>
</tr>
<tr>
<td>Student-Faculty Interaction</td>
<td>Student-Faculty Interaction</td>
</tr>
<tr>
<td></td>
<td>Effective Teaching Practices</td>
</tr>
<tr>
<td>Supportive Campus Environment</td>
<td>Quality of Interactions</td>
</tr>
<tr>
<td></td>
<td>Supportive Environment</td>
</tr>
</tbody>
</table>

*Note.* A description of the transition from NSSE 1.0 to NSSE 2.0, and the NSSE’s contents can be found in “NSSE: From benchmarks to engagement indicators and high-impact practices” (2014).

Table 2
*Learning outcomes measures by the NSSE*

<table>
<thead>
<tr>
<th>Academic and Interpersonal</th>
<th>Thinking critically and analytically</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Writing clearly and effectively</td>
</tr>
<tr>
<td></td>
<td>Working effectively with others</td>
</tr>
<tr>
<td></td>
<td>Speaking clearly and effectively</td>
</tr>
<tr>
<td></td>
<td>Acquiring job- or work-related knowledge and skills</td>
</tr>
<tr>
<td></td>
<td>Developing or clarifying a personal code of values and ethics</td>
</tr>
<tr>
<td>Application Gains</td>
<td>Solving complex real-world problems</td>
</tr>
<tr>
<td></td>
<td>Understanding people of other backgrounds</td>
</tr>
<tr>
<td></td>
<td>Analyzing numerical and statistical information</td>
</tr>
<tr>
<td></td>
<td>Being an informed and active citizen</td>
</tr>
</tbody>
</table>

Table 3
Original and adapted questions from the National Survey of Student Engagement.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>General NSSE</th>
<th>Quantitative Specific NSSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Higher-Order Learning</strong></td>
<td>During the current school year, how much has your coursework emphasized the following:</td>
<td></td>
</tr>
<tr>
<td>a)</td>
<td>Applying facts, theories, or methods to practical problems or new situations</td>
<td>a) Applying facts, theories, or methods to practical quantitative problems or new quantitative situations</td>
</tr>
<tr>
<td>b)</td>
<td>Analyzing an idea, experience, or line of reasoning in depth by examining its parts</td>
<td>b) Analyzing an idea, experience, or line of reasoning in depth by examining its quantitative parts</td>
</tr>
<tr>
<td>c)</td>
<td>Evaluating a point of view, decision, or information source</td>
<td>c) Using quantitative information to evaluate a point of view, decision, or information source</td>
</tr>
<tr>
<td>d)</td>
<td>Forming a new idea or understanding from various pieces of information</td>
<td>d) Forming a new idea or understanding from various pieces of quantitative information</td>
</tr>
<tr>
<td><strong>Reflective, &amp; Integrative Learning</strong></td>
<td>During the current school year, how often have you</td>
<td></td>
</tr>
<tr>
<td>a)</td>
<td>Combined ideas from different courses when completing assignments</td>
<td>a) Combined quantitative ideas from different courses when completing assignments</td>
</tr>
<tr>
<td>b)</td>
<td>Examined the strengths and weaknesses of your own views on a topic or issue</td>
<td>b) Examined the strengths and weaknesses of your own views on a quantitative topic or issue</td>
</tr>
<tr>
<td>c)</td>
<td>Connected ideas from your courses to your prior experiences and knowledge</td>
<td>c) Connected quantitative ideas from your courses to your prior experiences and knowledge</td>
</tr>
</tbody>
</table>
### Learning Strategies

During the current school year, how often have you

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>Identified key information from reading assignments</td>
<td>a)</td>
</tr>
<tr>
<td>b)</td>
<td>Reviewed your notes after class</td>
<td>b)</td>
</tr>
<tr>
<td>c)</td>
<td>Summarized what you learned in class or from course materials</td>
<td>c)</td>
</tr>
</tbody>
</table>

### Collaborative Learning

During the current school year, how often have you

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>Asked another student to help you understand course material</td>
<td>a)</td>
</tr>
<tr>
<td>b)</td>
<td>Explained course material to one or more students</td>
<td>b)</td>
</tr>
<tr>
<td>c)</td>
<td>Prepared for exams by discussing or working through course material with other students</td>
<td>c)</td>
</tr>
<tr>
<td>d)</td>
<td>Worked with other students on course projects or assignments</td>
<td>d)</td>
</tr>
</tbody>
</table>

### Student-Faculty Interaction

During the current school year, how often have you

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>Talked about career plans with a faculty member</td>
<td>a)</td>
</tr>
<tr>
<td>b)</td>
<td>Worked with a faculty member on activities other than coursework (committees, student groups, etc.)</td>
<td>b)</td>
</tr>
<tr>
<td>Effective Teaching Practices</td>
<td>Supportive Environment</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td>During the current school year, to what extent have your instructors done the following:</td>
<td>How much does your institution emphasize the following:</td>
<td></td>
</tr>
<tr>
<td>a) Clearly explained course goals and requirements</td>
<td>a) Providing support to help students succeed academically</td>
<td></td>
</tr>
<tr>
<td>b) Taught course sessions in an organized way</td>
<td>a) Using learning support services (tutoring services, writing center, etc.)</td>
<td></td>
</tr>
<tr>
<td>c) Used examples or illustrations to explain difficult points</td>
<td>b) Using quantitative learning support services (tutoring services, Math Lab, etc.)</td>
<td></td>
</tr>
<tr>
<td>d) Provided feedback on a draft or work in progress</td>
<td>d) Providing prompt and detailed feedback on tests or completed assignments requiring quantitative reasoning</td>
<td></td>
</tr>
<tr>
<td>e) Provided prompt and detailed feedback on tests or completed assignments</td>
<td>e) Provided prompt and detailed feedback on tests or completed assignments requiring quantitative reasoning</td>
<td></td>
</tr>
</tbody>
</table>

Items used with permission from The College Student Report, National Survey of Student Engagement, Copyright 2001-18 The Trustees of Indiana University.
Table 4
Demographics

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>214</td>
<td>(68.6%)</td>
</tr>
<tr>
<td>Male</td>
<td>94</td>
<td>(31.4%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>n</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian</td>
<td>4</td>
<td>(1.3%)</td>
</tr>
<tr>
<td>Asian</td>
<td>28</td>
<td>(8.9%)</td>
</tr>
<tr>
<td>Black</td>
<td>13</td>
<td>(4.2%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>25</td>
<td>(8.0%)</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>5</td>
<td>(1.6%)</td>
</tr>
<tr>
<td>White</td>
<td>265</td>
<td>(84.7%)</td>
</tr>
</tbody>
</table>

*Note.* Participants were able to select multiple ethnicities that they identified with.

Table 5
Perceived Quantitative Learning Gains

- How much has your experience at this institution contributed to your knowledge, skills, and personal development in the following area: (Very Much, Quite a Bit, Some, Very little)
  - a. Analyzing numerical and statistical information

*Note.* Items xx and xx used with permission from The College Student Report, National Survey of Student Engagement, Copyright 2001-18 The Trustees of Indiana University
Table 6  
*Natural World Self-Reported Learning Gain Scale*

How much has your experience at this institution contributed to your knowledge, skills, and personal development in the following areas:

a. Using graphical, symbolic, and numerical methods to analyze, organize, and interpret natural phenomena.

b. Discriminating between association and causation, and identifying the types of evidence used to establish causation

(Sundre, 2008)

Table 7  
*Quantitative Reasoning 9 Example Items*

1. Regarding the two graphical displays given below, which of the following statements is correct?

![Average Monthly Temperature Graphs](image)

a. Banebrook has the largest changes in temperature throughout the year.

b. Banebrook and Grove City temperatures exhibit exponential behavior throughout the year.

c. Neither of the above.

2. Suppose a researcher wants to test the hypothesis that exposure to cadmium in childhood causes neurological damage that reduces IQ. The researcher randomly selects 500 fourth graders, monitors their cadmium exposure for one year, and then tests each student’s IQ. The researcher finds that as cadmium exposure increases, IQ declines. Can the researcher conclude from the observed association between cadmium exposure and intelligence that cadmium causes reduced IQ?

a. No. The researcher did not include enough persons in the study.

b. No. There may be a third variable associated with exposure to cadmium that actually causes the lowered IQ.

c. Yes. The researcher followed the scientific method.

d. Yes. An association between the amount of cadmium exposure and lowered IQ is exactly what we would predict from the hypothesis.

*Note.* Example items taken from Hurney et al., 2011.
Table 8

*NSSE Indicators*

<table>
<thead>
<tr>
<th>Reflective &amp; Integrative Learning *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Strategies *</td>
</tr>
<tr>
<td>Quantitative Reasoning</td>
</tr>
<tr>
<td>Collaborative Learning *</td>
</tr>
<tr>
<td>Discussions with Diverse Others</td>
</tr>
<tr>
<td>Student-Faculty Interaction *</td>
</tr>
<tr>
<td>Effective Teaching Practices *</td>
</tr>
<tr>
<td>Quality of Interactions *</td>
</tr>
<tr>
<td>Supportive Environment *</td>
</tr>
<tr>
<td>Campus Environment</td>
</tr>
</tbody>
</table>

*Note.* * indicates that the subscale was used in both measures of student engagement.

Table 9

*Example NSSE General Engagement Items*

<table>
<thead>
<tr>
<th>During the current school year, how often have you done the following?</th>
</tr>
</thead>
<tbody>
<tr>
<td>o Summarized what you learned in class or from course materials</td>
</tr>
<tr>
<td>o Worked with other students on course projects or assignments</td>
</tr>
<tr>
<td>o Discussed course topics, ideas, or concepts with a faculty member outside of class</td>
</tr>
</tbody>
</table>

(NSSE: Engagement indicators, 2017)

Table 10

*Example NSSE Specific Engagement Items*

<table>
<thead>
<tr>
<th>During the current school year, how often have you done the following?</th>
</tr>
</thead>
<tbody>
<tr>
<td>o Summarized what you learned in a Cluster 3 class or from quantitative course materials</td>
</tr>
<tr>
<td>o Worked with other students on course projects or assignments involving quantitative content</td>
</tr>
<tr>
<td>o Discussed quantitative course topics, ideas, or concepts with a faculty member outside of class</td>
</tr>
</tbody>
</table>
Table 11

*Student Opinion Scale: Example Items*

1. I engaged in good effort throughout this test
2. While taking these examinations, I could have worked harder on them *
3. While taking these tests I was able to persist to completion of the task

*Note.* *Indicates reverse scoring (Thelk, Sundre, Horst, & Finney, 2009)*

Table 12

*Table comparing students QR-9 gains across responses levels to NSSE SRLG*

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean Gain</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Little (0)</td>
<td>85</td>
<td>1.424</td>
<td>3.822</td>
</tr>
<tr>
<td>Some (20)</td>
<td>123</td>
<td>1.203</td>
<td>3.637</td>
</tr>
<tr>
<td>Quite a bit (40)</td>
<td>83</td>
<td>0.121</td>
<td>3.362</td>
</tr>
<tr>
<td>Very much (60)</td>
<td>22</td>
<td>0.636</td>
<td>3.513</td>
</tr>
</tbody>
</table>
Table 13

*Table comparing students QR-9 gains across responses levels to specific SRLG*

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean Gain</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Little (0)</td>
<td>79</td>
<td>0.9747</td>
<td>4.291</td>
</tr>
<tr>
<td>Some (20)</td>
<td>95</td>
<td>1.4000</td>
<td>3.204</td>
</tr>
<tr>
<td>Quite a bit (40)</td>
<td>110</td>
<td>0.6273</td>
<td>3.619</td>
</tr>
<tr>
<td>Very much (60)</td>
<td>29</td>
<td>0.4828</td>
<td>2.972</td>
</tr>
</tbody>
</table>

Table 14

*Correlations between engagement and measures of learning on general engagement*

<table>
<thead>
<tr>
<th>1 General Engagement</th>
<th>2 General Self-reported Learning</th>
<th>3 Specific Self-reported Learning</th>
<th>4 Direct Measured Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>.376**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.520**</td>
<td>.632**</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-.036</td>
<td>-.146*</td>
<td>-.092</td>
</tr>
</tbody>
</table>

*Note. *p<.05, **p<.001*
Table 15
Correlations between engagement and measures of learning on QR specific engagement

<table>
<thead>
<tr>
<th></th>
<th>1 Specific Engagement</th>
<th>2 General Self-reported Learning</th>
<th>3 Specific Self-reported Learning</th>
<th>4 Direct Measured Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.489**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.586**</td>
<td>.732**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.009</td>
<td>-.077*</td>
<td>.011</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. *p<.05, **p<.001
Figure 1. Diagram illustrating student learning as influenced by indirect factors.
<table>
<thead>
<tr>
<th>Good Content + Good Teaching + Good Opportunities</th>
<th>Poor Content + Poor Teaching + Poor Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Natalie (Engaged)</strong></td>
<td><strong>Irwin (Disengaged)</strong></td>
</tr>
<tr>
<td>Substantial Learning</td>
<td>Moderate Learning</td>
</tr>
<tr>
<td>Moderate Learning</td>
<td>Minimal Learning</td>
</tr>
</tbody>
</table>

*Figure 2.* Institutional factors and student engagement influence on learning.
References


CESQ, 2007 http://cseq.indiana.edu/cseq_generalinfo.cfm


Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we?. *Psychological bulletin, 74*(1), 68.


http://doi.org/10.1016/S0099-1333(02)00339-7


ETS 2017 https://www.ets.org/heighten/scores/


http://doi.org/10.1080/0142159022020687


http://doi.org/10.1080/00091383.2013.786985


NSSE: From benchmarks to engagement indicators and high-impact practices, (2014).


http://doi.org/10.1080/00091380903449060


http://doi.org/10.1023/A:1018774311468

http://doi.org/10.1353/csd.2006.0061

http://doi.org/10.1007/s11162-005-8885-3


