

ESTABLISHING NORMATIVE BENCHMARKS FOR ON-TASK, OFF-TASK, AND
DISRUPTIVE BEHAVIORS IN EARLY ELEMENTARY CLASSROOMS

Elizabeth Holland

A dissertation

submitted in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY

University of Washington

2015

Reading Committee:

Clayton Cook, Chair
Jim Mazza
Aaron Lyon

Program Authorized to Offer Degree:

College of Education

©Copyright 2015
Elizabeth Holland

University of Washington

Abstract

Establishing Normative Benchmarks for On-Task, Off-Task, and Disruptive Behaviors in
Early Elementary Classrooms

Elizabeth Holland

Chair of Supervisory Committee:

Assistant Professor Clayton Cook, Ph.D.

College of Education

Systematic direct observations of on-task, off-task, and disruptive behaviors were conducted in Kindergarten, First, and Second grade classrooms in order to establish normative benchmarks for academic engaged time for early elementary students at the individual, school, and district level. Descriptive statistics of the participating students ($n=6,592$) indicated the overall mean academic engaged time in early elementary classrooms is approximately 80%. Visual analysis of the variability of academic engagement suggested that there is variability evident within districts and schools. Finally, multiple linear and logistic regressions were performed to determine associations between school-level variables and two outcome variables: Academic Engaged Time and Disruptive Behavior. Results indicated that the school-level demographic variables of Free and Reduced Lunch Status (a school-level socioeconomic measure) and school

percentage of white/non-Hispanic students were significantly associated with school-wide academic engagement. The second outcome variable of disruptive behavior was significantly associated with school academic achievement as measured by reading and math curriculum-based measurements. Future implications about data based decision-making, intervention planning, and monitoring of school and classroom wide behavioral management using the systematic observational data is discussed.

Table of Contents

Chapter 1: Introduction

Problem Description 1

Purpose of Dissertation 4

Research Questions 5

Chapter 2: Literature Review

Overview of Student Engagement in a Response to Intervention Approach..... 8

Defining Response to Intervention/Multi-Tiered Systems of Support 9

Essential Components of RtI/MTSS 10

 Evidence Based Practices and RtI/MTSS 10

 Defining Evidence-Based Practices 11

 Universal Screeners and Progress Monitoring..... 13

 Problem Solving Model 16

 Behavioral and Social-Emotional RtI/MTSS..... 17

 Academic and Behavioral RtI/MTSS Improves Student Engagement .. 24

Student Engagement 25

 Student Engagement and Academic Outcomes 25

 Defining Student Engagement 26

 Student Engagement and Construct Validation 30

 Measuring Student Engagement 32

 Simple View of Learning and Academic Engagement..... 33

Systematic Direct Observations 34

Importance of Establishing Behavioral Benchmarks.....	37
Purpose of Dissertation Study.....	41
Chapter 3: Methods	
Setting and Participants.....	44
Design	44
Measures and Procedures.....	45
Systematic Direct Observations.....	45
Curriculum-Based Measurements.....	48
School Level Data.....	52
Data Analytic Strategies and Research Questions	52
Chapter 4: Results	
Research Question #1	58
Research Question #2	63
Research Question #3	64
Chapter 5: Discussion	
General Discussion	66
Limitations	69
Future Implications	71
Conclusion	73
References.....	75
Appendix.....	89

CHAPTER 1: INTRODUCTION

Student engagement in the classroom contributes to positive outcomes for students, including improved academic performance and social and cognitive development (Finn, 1993; Newman, 1992). Further, research has consistently demonstrated that student *disengagement* has a strong relationship to low academic performance, higher school drop out rates, and problematic behaviors (Finn, 1989). Given the many positive outcomes associated with student engagement, educators and researchers aim to increase student engagement as it has been shown to be an integral component of learning in the classroom (Rathvon, 1999). When students are not engaged in the classroom, a multitude of negative outcomes occur because disengagement negatively affects academic and behavioral functioning, whereas student engagement increases academic achievement, school completion, and pro-social behaviors that contribute to long-term positive outcomes.

In recent years, the United States educational system has taken a systematic approach in the delivery of educational services known as Response to Intervention (RtI) or Multi-Tiered Systems of Support (MTSS). RtI/MTSS is a proactive, data-driven approach that increases positive student outcomes by implementing evidence-based instruction for all students. One integral component of RtI/MTSS is that it is responsive, meaning that instead of waiting for a student to struggle academically and/or behaviorally, educators actively pursue implementing academic and behavioral strategies that have been shown to improve student performance. This population-based approach is systematic and depends on empirical evidence of ‘what works’ in education. Since student engagement has been shown to be imperative in producing positive student

outcomes, an effective RtI/MTSS approach should aim to increase student engagement at all service delivery levels.

A RtI/MTSS approach in education requires schools to actively and adequately monitor students' performance because data-based decision making is an integral part of ensuring that students are responding to their educational programming (Shapiro, Hilt-Panahon, & Gischlar, 2010). Because student engagement is incompatible with problem behavior and has been shown to increase academic achievement, educators and researchers should increase student engagement for all students *before* problems are identified. This requires a population-based approach where student engagement is operationally defined, measured, and interpreted through psychometrically sound normative behavioral data. Thus, student behavior should be regularly monitored in order to ensure that educators are proactive in their approach and are able to interpret student behavior within multiple contexts. Normative data of student behavior allows educators to better measure and interpret individual student behavior, overall classroom behavior at different grade levels, and school-wide behavioral functioning.

School Psychologists often rely on measures of student behavior in order to provide information about an individual's functioning within the classroom (Hintze, Volpe, & Shapiro, 2002). In fact, the Individuals with Disabilities Improvement Act (IDEIA, 2004) requires technically adequate tools to assess and monitor students' behavior in schools. Assessing and monitoring students' behavior in the classroom is an essential component in an RtI/MTSS educational delivery system (Sugai & Horner, 2002). Data-based decision making in schools occurs not only at the universal level where *all* student performance is measured (i.e., standardized state assessments), but

student response to intervention(s) to more intensive remediation should be actively monitored so that data-based decision making can occur at all performance levels. In recent years, US schools have focused on implementing a RtI/MTSS approach in their educational delivery system, but much of this focus has been on academic data while adequate behavioral data continues to be an area of need in schools since there is far less research in regards to the reliable measurement of student behavior in the classroom (Briesh, Chafouleas, & Riley-Tillman, 2010). This is surprising given that schools are expected to implement both effective academic and behavioral instruction and interventions to students, which relies on normative data to interpret student academic and behavioral functioning.

One of the most often utilized and cited ways to measure student behavior are behavioral observations. Systematic direct observational (SDO) systems have been shown to be valid measures of behavior and can yield valuable information for monitoring the effectiveness of an intervention and intervention planning (Sattler & Hoge, 2006). SDOs provide quantitative data through time sampling pre-determined operationally defined behaviors. For example, the Behavior Observation of Students in Schools (BOSS) developed by Shapiro (2004) is one of the most commonly used SDO tools used by school psychologists. SDOs provide quantitative data of student behavioral engagement that can provide baseline information about a student's current behavioral functioning and SDOs are effective in monitoring a student's response to intervention. Further, SDOs are cited as being more objective in nature than compared to teacher subjective judgment in determining the nature and severity of problem behaviors (Shapiro, 2010). School psychologists often utilize SDOs when they are conducting

individual evaluations for students suspected of having a disability. In order to interpret a particular student's behavioral data, SDOs use local normative comparison by using peers in the classroom of the same gender. This peer comparison utilizes local norms within the classroom to interpret a student's behavior. When SDOs are conducted in classrooms, their purpose is to measure a student already exhibiting difficulties; they are not utilized to measure engagement before problems have occurred and peer comparison data exists only within the context of a particular classroom.

In a review of research, little is known about behavioral trends in classrooms. In fact, there is no known SDO quantitative data that measures behavioral student engagement beyond local norms. Given that a RtI/MTSS system in education requires that data-based decision making occur at all service delivery levels, it is suggested that more research is needed to better understand student behavior within the classroom. Such behavioral engagement metrics could provide educators and researchers valuable information that would help interpret student engagement at the individual, classroom, school, and district levels. Just as educators understand national and local norms of reading expectations at early elementary grades, it is proposed that an additional lens for understanding behavioral engagement trends should be established.

Purpose of this Dissertation Study

In this study, systematic direct observations of on-task, off-task, and disruptive behaviors were conducted in Kindergarten, First, and Second grade classrooms. Descriptive statistics of the participating students ($n=6,592$) were analyzed in order to further contribute to the research on behavioral data of students in early elementary classrooms. Specifically, descriptive statistics of students' classroom behaviors will assist

researchers to better interpret student classroom behavior and can be used to further understand student engagement in the classroom. In addition, descriptive statistics of behavioral data can be used as a tool to better understand the variability in student classroom behavior amongst students and within classrooms, schools, and districts. This study aimed to add to the body of literature on the validity of SDOs as a reliable measure of student behavior by correlating student behavioral engagement data with academic performance data. Further, implications about SDO data and its utility for data-based decision making within a RtI/MTSS delivery system, classroom management, intervention planning, and monitoring of student engagement at the student, classroom, school, and district levels will be discussed. Finally, school-level variables, such as free and reduced lunch status, student demographics, and academic performance will be examined to better understand what school-level factors best predict behavioral engagement. Specifically, this dissertation proposal aimed to answer the following questions:

Research Question #1: What is the Kindergarten, First, and Second grade normative data for Academic Engaged Time (AET) and Disruptive Behavior (DB)?

Research Question #2: What is the variability in normative data across classrooms, schools, and districts?

Research Question #3: To what extent are school-level variables (percent of students qualifying for Free and Reduced Lunch [FRL], percent of white/non-Hispanic students, and academic achievement data) associated with school-level norms regarding Academic Engaged Time (AET) and Disruptive Behavior (DB)?

CHAPTER 2: LITERATURE REVIEW

The review of the literature that follows is intended to provide adequate coverage of the background research that builds the case for and significance of this dissertation study. The chapter is organized according to the following. First, a general discussion about RtI/MTSS is provided, which will include an overview of how RtI/MTSS is defined and the essential components of an RtI/MTSS service delivery model with a focus on how evidence-based interventions are conceptualized in education. This section will also review research on behavioral and mental health RtI/MTSS and the current state of behavioral evidence-based practices in K-12 education. Second, the construct of student engagement will be discussed, which will include past and current conceptualizations of engagement and how it is measured in educational settings. Both affective and behavioral engagement will be thoroughly reviewed and how student engagement is defined in this study will be discussed. Third, SDOs will be defined and how they measure student engagement. This section will also explore the relevance of SDO data in an RtI/MTSS model. Finally, a discussion of the importance of having norms for behavioral student engagement and the current limited research on behavioral engagement will be discussed. Throughout this literature review specific gaps in the extant research will be highlighted to identify needs for further research. The end goal is to build the significance for the present work of establishing normative benchmarks of student engagement in early elementary classrooms.

Overview of Student Engagement in RtI/MTSS Approach

One aim of RtI/MTSS is that it provides quality general education instruction for *all* students in order to ensure that students receive proactive, evidence-based instruction that emphasizes early intervention. This systematic service delivery model ensures that students have quality environments in their classrooms, which is essential for learning to occur (Burns & Wagner, 2008; Soukup, Wehmeyer, Bashinksi, & Bovaird, 2007). In fact, student engagement has long been cited as an effective component of a classroom environment that increases learning (Reith & Evertson, 1988). Given the close relationship between student achievement and student engagement, US schools have made efforts in recent years to increase student engagement (Marks, 2000). For example, some schools have created small schools within schools, increased social supports in schools (e.g., multi-year advisory groups, staying with the same teacher and class for two years), focused on active involvement in learning (e.g., project-based learning, increasing challenge), and made efforts to increase parental and student involvement in curricula and evaluations (e.g., math nights, parent conferences, student-led conferences) (Marks, 2000). Clearly, increasing student engagement is an important goal in K-12 education and many efforts have been made to increase student engagement at the student, classroom, school, and community levels. However, despite the importance of student engagement, there has been relatively little research about student engagement and how student engagement should be conceptualized and measured within an RtI/MTSS model.

Defining Response to Intervention (RtI) or Multi-Tiered System of Support (MTSS)

Serving students in a multi-tiered service delivery model, known in education as RtI/MTSS, has been at the forefront in educational reform in the last decade. In fact, the Individuals with Disabilities Improvement Act (IDEIA, 2004) called on K-12 educators to use scientifically based practices in public education using an RtI/MTSS model. This movement in school-based delivery services adopts a public health model in which evidence based practices (EBPs) are disseminated at all levels of service delivery, which range from primary prevention practices that are universally administered to tertiary prevention practices that are intensive interventions for an individual (Walker & Shinn, 2010). RtI/MTSS is conceptualized as three-tiers where there are universal supports for “most” children (approximately 80%), targeted supports for “some” children (approximately 15%), and intensive supports for “few” children (approximately 5%).

Brown-Chidsey and Steege (2005) assert that there are two essential components of RtI/MTSS that differentiate it from other educational practices: *systematic* and *data-based*. *Systematic* refers to the whole system approach to educational service delivery; for example, RtI/MTSS is not a curriculum or an instructional strategy, but instead is a systems approach to serving all students, which involves general instruction and assessment (all students), supplementary instruction and assessment (some students), and specialized instruction and assessment (a few students). *Data-based* refers to the decision making process in a RtI/MTSS model, where response to intervention is benchmarked and progress monitored throughout the school year; data-based decision making occurs at all levels in a RtI/MTSS model (Brown-Chidsey & Steege, 2005). Through a system-wide approach that utilizes data-based decision making, RtI/MTSS aims to improve

student outcomes and to identify students who may require more intensive intervention. Gresham (1991) succinctly writes that RtI/MTSS is defined as the change in behavior or performance as a function of an intervention. RtI/MTSS is proactive, meaning that it attempts to serve students before they struggle or fail in school. This is similar to a medical model where we know that prevention is paramount in improving outcomes. This public health model takes on a population-based approach, which prioritizes prevention practices.

Essential Components of RtI/MTSS

EBPs in RtI/MTSS Framework. A central component of RtI/MTSS is the delivery of EBPs across each tier (Walker & Shinn, 2010). In fact, one of the requirements under federal law is that schools should provide scientific, research-based interventions at all levels of education (i.e., general education, secondary supports, and special education) before determining if a child has a learning disability (IDEIA, 2004). Thus, EBPs should be integrated at all levels in educational service delivery, which creates a proactive system where EBPs are prevalent throughout the educational system. Becker and Domitrovich (2011) argue that integrating EBPs in schools at each service level allows school staff to more efficiently and meaningfully deliver EBPs because it increases a common language, conceptual framework, practice opportunities, and consistency across students and staff. Further, Becker and Domitrovich (2011) suggest that an integrated approach prevents the delivery of multiple uncoordinated programs where students cannot move seamlessly through each service delivery level. In summary, a core component to a successful RtI/MTSS service delivery model in schools is to implement EBPs at all levels of education. In fact much of the increased focus on EBP

research in education is a result of the changing landscape in how educational services are delivered in public education. Many have argued that this proactive approach where EBPs are delivered in general education settings avoids the current pervasive “wait to fail” model in education where EBPs are only sought out when problems are already identified.

Defining Evidence Based Practices (EBPs). EBPs and programs are broadly defined as scientific knowledge about programs, practices, interventions, or treatments that have shown to have meaningful effects on student outcomes (Hoagwood, 2003-04). Federal legislation such as the No Child Left Behind Act of 2001 (NCLB, 2001) mandated that educators should use “scientifically based research” to determine which practices and instruction should be implemented. Further, the Individuals with Disabilities Education Improvement Act of 2004 (IDEIA, 2004) reiterated the need to utilize scientifically based interventions to improve student outcomes and stated the need to better identify rigorous, systematic studies in educational research. The study of EBPs originally emerged in the field of medicine when medical professionals in the 1970s began to demand more research-based methods and practices (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). The educational field has followed the medical model in establishing the need to identify and implement EBPs. In fact, much of the push for EBPs in schools has been a result of a long standing history of adopting non-evidence based practices or fads in education that were found not to be effective and were often costly (e.g., whole language approach to reading instruction, DARE program) (Whitehurst, 2001). Thus, organizations such as the What Works Clearing House, the National Professional Development Center on Autism Spectrum Disorder, and the Promising

Practices Network were established and are all examples of the increased focus and research on developing and disseminating EBPs in education. Despite federal legislation and increasing acknowledgment of the need for EBPs in education, many educators do not have the necessary knowledge to determine what constitutes an EBP (NCEE, 2003). Further, EBPs are differentially emphasized, identified, and implemented in general education versus special education settings (Cook & Schirmer, 2006). Finally, mental and behavioral EBPs are considered to be lacking in educational services according to federal institutions (President's Freedom Commission, 2003).

EBPs have been defined and conceptualized in different ways. Tankersley, Harjusola-Webb, and Landrum (2008) assert that an evidence-based practice “refers to instructional strategies or educational programs shown to produce consistent positive student outcomes” (p.83). Whitehurst (2001) defined EBPs in similar terms citing that empirical evidence must be used in determining an EBP; however, Whitehurst (2001) expanded on this definition by stating that EBPs are an “integration of professional wisdom with the best available empirical evidence” (p. 3). Cook and Odom’s (2013) conceptualization of EBPs focuses on the research standards needed to determine an EBP and stated that EBPs are both practices and programs shown to yield meaningful effects as demonstrated by high-quality research. Further, Cook and Cook (2011) state that EBPs are different from previous approaches in education because in the past educators have instituted *best practices*, whereas EBPs are *research-based practices* that have been found to be effective under “prescribed, rigorous” research standards. A common theme among all these definitions and descriptions of EBPs is that the identification of EBPs requires rigorous empirical scientific research. However, within educational scholarship,

there are a myriad of theoretical orientations that range from philosophical researchers to empirical researchers. While qualitative and quantitative research coexist within educational research, much of the EBP research resides within an empirical orientation that views efficacy through measureable outcomes.

A major component of determining EBPs across the many definitions described by researchers is the need for rigorous research standards that demonstrate *quality* and *quantity* of evidence. The National Center for Education Evaluation and Regional Assistance (NCEE, 2003) provided a guide for educators to better understand and determine which practices and programs could be considered evidence-based. Their guide notes that randomized control trials are the “gold standard” in research that can best identify an EBP. A randomized control trial includes a treatment and control group where effects are measured to help determine if an intervention causes an outcome in which subjects are randomly assigned into either the treatment or control conditions. A randomized control trial that is well designed and implemented contributes to the *quality* of research and the *quantity* of evidence can be demonstrated in the amount of studies where EBPs are shown to be effective in “typical school settings” (NCEE, 2003).

Universal Screeners and Progress Monitoring. In order to identify at-risk students in an RtI/MTSS framework, quick screeners are given to all students periodically throughout the school year (Fuchs & Fuchs, 2006; Hosp et al., 2007). Screeners not only identify at-risk students, they also are an effective method to monitor the overall general education instruction (Cook, Browning-Wright, Gresham, & Burns, 2010). For example, if a teacher screened their class and found that 40% of students were off-task, then the teacher would want to adjust classroom management practices for the

entire classroom as it would show that it is a general education instruction problem, not a “within child” problem. Thus, an RtI/MTSS approach that utilizes universal screeners and progress monitoring tools improves teaching as well as student outcomes.

Another main tenet of RtI/MTSS is monitoring student progress to interventions (Cook et al., 2010). This ensures that students who are receiving supplemental instruction (Tier 2) or more intensive interventions (Tier 3) are responding positively. Student progress is monitored consistently in order to ensure fidelity of treatment and it assists educators in making data-based decisions about the kind of instruction and the amount and intensity of interventions needed. In many current traditional models of education, there are too many students qualifying for special education that have not had the benefit of EBPs, which is often a result of poor universal screening and progress monitoring (Gresham, Reschly, & Shinn, 2010). Unfortunately, too many students are being labeled as learning or emotional-behaviorally disabled when in fact they would be better served in an RtI/MTSS framework where their deficits are addressed early and effectively.

In the current data-driven educational environment, schools are required to monitor students’ progress over time and essential components of an RtI/MTSS model include universal screening and progress monitoring. Data-based decision making is a core feature of a problem-solving model that promotes prevention and intervention within a multi-tiered service delivery model. According to Power, Mautone, and Ginsburg-Block (2010), universal screening is defined as a preventative strategy that collects data on *all* students. This universal approach allows for early intervention and prevention (Ikeda, Neesen, & Witt, 2008). Screening data not only provides information about individual students, but also is a useful tool to evaluate classroom, school building, and

district performance (Power et al., 2010). According to Sprague, Cook, Browning-Wright, and Sadler (2008) universally screening the entire school population should typically occur three times per year in the Fall, Winter, and Spring. Cook et al. (2011) noted that universal screeners possess three main characteristics: 1. A universal screener should be reliable, so they accurately predict performance over time, 2. A universal screener should be valid, so that they correctly identify student performance, and 3. A universal screener should be “cost beneficial” in that they are easy to administer (time cost beneficial) and are feasible to administer within schools. While universal screening is an imperative component in an RtI/MTSS system, there is far more research on universal academic screeners than compared to universal behavioral screeners.

Progress monitoring tools are similar to universal screeners in that they are reliable, valid, and cost beneficial. A key difference of progress monitoring tools when compared to universal screeners is that it is employed to evaluate the effects of intervention (Power et al., 2010). According the U.S. Department of Education, progress monitoring serves different functions within a multi-tiered service delivery model. For students and classrooms identified as at-risk on universal screeners, progress monitoring occurs more frequently and can occur 1-2 times weekly or monthly (Power et al., 2010). Progress monitoring tools for academic achievement are widely available and implemented in schools, but progress monitoring of behavior is less developed (Cook et al., 2011). In fact, behavioral RtI/MTSS is a burgeoning area within educational research and more quantitative data about behavioral trends in K-12 education is needed in order to better understand, screen, and progress monitor students’ behavior in schools.

Problem Solving Model. Universal screening and progress monitoring by their very nature contribute to a problem-solving model in education. As previously described, RtI/MTSS takes a system wide public health model approach that at its very core increases preventative practices to increase positive student outcomes. Deno (2005) outlined five essential components in a RtI/MTSS system: 1. Problem Identification, 2. Problem Definition, 3. Designing Intervention Plans, 4. Implementing the Intervention, and 5. Problem Solution. This systematic delivery service through these five steps guides educational decisions that affect students at all levels in an RtI/MTSS model. Deno (2005) breaks down the five essential components of RtI/MTSS delivery into assessment procedures and evaluation decisions. Specifically, in step one, *Problem Identification*, Deno (2005) asserts that all students' academic and behavioral performance should be assessed (i.e., universal screening) and educators should use this data to ask themselves: Does a problem exist? In the second step, *Problem Identification*, educators use assessment procedures to quantify the discrepancy of the performance deficit (e.g., gap analysis) and answer the question: Is the problem important (i.e., significant)? (Deno, 2005; Brown-Chidsey, 2007). In the third step, *Designing Intervention Plans*, educators develop a plan of action through goal planning and solution hypotheses and answer the question: What is the best solution based on what we know? (Deno, 2005; Steege & Brown-Chidsey, 2007; Cook et al., 2010). Next, educators *Implement the Intervention*, which entails monitoring the fidelity of the intervention and progress monitoring the response to the intervention and answers the question: Is the intervention being implemented and progressing as planned? (Deno, 2005). Finally, in the fifth step, *Problem Solution*, educators reexamine the discrepancy in student performance and

answer the question: Is the problem being solved through this intervention? (Deno, 2005; Brown-Chidsey, 2007). This problem-solving model guides educators to make data-based decisions. It allows for meaningful decision-making where EBPs can be implemented across tiers. Cook et al. (2010) succinctly writes that a problem solving model in a RtI/MTSS service delivery model helps educators decide if they should maintain existing supports, modify existing supports, decrease supports (i.e., lower down a tier), or increase supports (i.e., bump up a tier). While problem identification and problem analysis has been well established in providing academic interventions at all levels in an RtI/MTSS system, there is a need to better establish this problem-solving model in behavioral interventions. For example, normative benchmarks in reading, writing, and math have been well established in curriculum-based measurements to better understand a student's academic performance and this normative data allows educators to make data-based decision making when analyzing the discrepancy in academic performance compared to national and local norms. However, there is a lack of normative behavioral data at the national and local level, which could assist in the decision making process in a behavioral RtI/MTSS model.

Behavioral and Social-Emotional RtI/MTSS. Historically, behavioral management in schools has been *reactive* as opposed to *proactive* (Sprick & Borgmeier, 2010). Reactive methods include disciplinary methods such as detention, suspension, or public reprimands. Sprick and Borgmeier (2010) also note that behavioral management in schools are also often *exclusionary* in that they exclude misbehaving students from general education (e.g., expulsion, in-school and out-of-school suspensions). Despite the fact that punitive and reactive discipline has been shown to be largely ineffective, many

schools continue to solely use reactive methods for behavioral management in schools (Sprick, Knight, Reinke, & McKale, 2007). As Sprick and Borgmeier (2010) write, “Punishment has rarely changed student behavior in school, but the severe consequences sometimes delivered to students have inadvertently encouraged them to leave the educational system sooner, either by being expelled or by voluntarily dropping out” (p. 437). This is especially alarming given the disheartening high school graduation rates in the United States (approximately 25-30% of high school students drop out) (Sprick & Borgmeier, 2010). Further, Sugai and Horner (2002) noted that reactive systems in schools are ineffective and do not create positive school climates.

The US Public Health Service and the American Academy of Pediatrics have recognized the harmful effects of untreated mental and behavioral health problems. Specifically, these organizations purport that untreated behavioral and mental health difficulties are associated with high-risk behaviors, attention problems, learning difficulties, increased violence, a higher occurrence of suicide, and mood disorders (DeSocio & Hootman, 2004). Mental and behavioral health difficulties have been shown to put children at-risk in schools for absenteeism, discipline problems, school retention, poor grades, school dropout, and juvenile delinquency (Doll & Cummings, 2008). In addition, Adelman and Taylor (2002) assert that student participation and engagement are prerequisites to accessing high quality instruction, suggesting that behavioral health problems such as disruptive behaviors and ADHD can have negative effects on a child’s ability to receive instruction and thus function academically. Clearly, children with behavioral and mental health difficulties that are left untreated result in numerous long-term negative outcomes. The National Advisory of Mental Health Council’s Workgroup

on Child and Adolescent Mental Health (2003) echoed this sentiment when they stated: “The extent, severity, and far-reaching consequences of mental health problems in children and adolescents make it imperative that our nation adopt a comprehensive, systematic, public health approach to improving the mental status of children.” In addition to the high personal costs to children with mental and behavioral health needs and their families, society costs in terms of crime, violence, and economics are great.

Given the great impact of mental and behavioral health difficulties on student outcomes and society at large, implementing behavioral supports using an RtI/MTSS approach is imperative in implementing a proactive and preventative approach in behavioral management in schools. Over the last few years in education, positive behavior support (PBS) in schools, which uses an RtI/MTSS approach in providing behavioral and social-emotional supports to all students, has become more prevalent. Sugai and Horner (2002) outline four elements of school-wide PBS: outcomes, research-validated practices, data-based decision making, and systems (e.g., routines, school processes, administrative support, program structures). Frey, Lingo, and Nelson (2010) assert that PBS has distinctive features, which include the following: an emphasis on prevention of problem behaviors, instruction in behavioral skills, a continuum of consequences for problem behaviors, functional-based interventions for children with the most challenging behaviors, a systematic approach to behavioral interventions, and data-based decision making that ensures fidelity and effectiveness of the interventions being implemented (p. 399). In a behavioral RtI/MTSS model that utilizes a proactive PBS approach, data-based decision making in schools has most often been measured using office discipline referral data (ODR) (Sugai & Horner, 2002). There remains a need to

have more data and tools available to assess behavior at each level in an RtI/MTSS model.

In addition to PBS in schools, an increased focus on school mental health services (SMHS) in K-12 public education in an RtI/MTSS model has been at the forefront of educational research in recent years. The Institute of Medicine (1988) outlined ten essential components in a population-based (i.e., RtI/MTSS) approach for public health services. Doll and Cummings (2008) adapted the Institute of Medicine's model by creating ten essential SMHS components in an RtI/MTSS system:

1. Monitoring of emotional, academic, behavior, and social skills
2. Assessment and analysis of mental health problems
3. Psychoeducation around mental health to children and their families
4. Collaboration within school, community, and family relationships
5. Develop administrative policies and plans that support SMHS
6. Implement policies and plans that support SMHS
7. Designate children and families to universal, targeted, and tertiary services as needed
8. School training on SMHS and tools to provided progress monitoring for students' responses to services
9. Assessment of the quality and availability of SMHS provided to children
10. Continued research on school mental health EBPs

Taking an RtI/MTSS approach to SMHS using these 10 essential components, allows for a systematic approach. Doll and Cummings (2008) assert that this population-based approach aims to create psychological well-being for all students that is critical for

student success, provides nurturing environments that foster children to “overcome minor risks and challenges,” supplies protective supports for higher risk groups of children, and remediates emotional, behavioral, and social difficulties (p.3).

The National Association of School Psychologists (NASP) (2006) states that mental health is not merely the absence of mental illness, but also includes social, emotional, and behavioral health that helps children cope in life. In fact, the American Psychological Association (APA) recommends a systems or ecological approach in order to improve child outcomes and increase effectiveness of children and adolescents’ behavioral and mental health services (Kratochwill, Hoagwood, Kazak, Weisz, Hood, Vargas, & Banez, 2012). Thus, a population-based approach to behavior and SMHS has been cited as an essential component for promoting mental health well-being, which is also preventative and proactive. One of the benefits to a universal approach is that it destigmatizes behavioral and mental health services (Seeley, Rohde, & Backen-Jones, 2010). Further, Doll and Cummings (2008) stated, “When school mental health services are framed around population-based strategies, they can be more fully integrated into the core activities of the schools” (p. 2). In a meta-analysis of 213 school based universal social-emotional learning programs designed to promote student behavioral and mental health, Durlak, Weissberg, Dymniki, Taylor, and Schellinger (2011) found that social emotional learning and behavioral supports in grades K-12 led to increased academic performance, social-emotional skills, and positive attitudes and behaviors.

Many universal school behavioral and mental health practices include group taught curriculum in classrooms that are often part of a manualized program using a variety of instructional methods (Seeley et al., 2010). In a review of behavioral and

school mental health universal supports, Seeley et al. (2010) found that preventative programs were primarily based on cognitive-behavioral therapy, interpersonal therapy, problem solving skills training, and social skills training approaches (p.374). Cook et al. (2011) provided an overview of typical Tier 1 universal supports found in schools, including school-wide positive behavior supports (PBS), proactive classroom management strategies for general education teachers, social skills or social-emotional general education learning curricula, and the good behavior game, which is a class-wide behavior management instructional tool. Typical Tier 2 behavior supports include EBPs such as behavior contracts, self-monitoring strategies, school-home note interventions, mentor programs (such as Check-in/Check-out), behavioral coaching, and peer interventions that use positive reporting from peers to increase acceptance and improve peer ecology (Cook et al., 2011). Sugai and Horner (2002) described Tier 3 behavioral and social emotional EBPs as being necessary for students who display “clear” symptoms of social, emotional, and behavioral problems and should represent about 1-5% of the school population. Behavioral and SMHS for Tier 3 often refers to one-to-one therapy where treatment goals are highly individualized. These interventions are more rigorous and can include interventions such as cognitive-behavioral therapy, family counseling, parent training, case management, and home visits (Christenson, Whitehouse, & VanGetson, 2008). Often children demonstrating intensive levels of support require highly individualized assessments, such as a functional behavior assessment (FBA) in order to identify the function(s) to problem behavior(s) and create a behavior intervention plan (BIP) that is highly individualized and based on an assessment of antecedents, behaviors, and consequences. Progress monitoring of response to interventions at Tier 3

are more closely monitored when compared to Tier 1 and Tier 2. Some educators view restricted settings, such as emotional behavioral disturbed (EBD) settings as Tier 3 services, while others consider such restrictive settings as beyond Tier 3.

Screening of behavioral and mental health problems is an essential component in an RtI/MTSS behavioral model. The importance of identifying behavioral and mental health needs and improving services through universal screeners was recognized by the President's New Freedom Commission on Mental Health (2003), Individual with Disabilities Education Improvement Act (IDEIA, 2004), the No Child Left Behind Act (NCLB, 2002), and US Department of Education (2002). These federal institutions and legislation asserted that screening all students for behavioral and mental health needs is essential for the prevention of problems and identification of service needs. In fact, IDEIA (2004) noted that up to 15% of special education funding could be used for the early screening, prevention, and intervention in order to decrease referrals to special education services (Walker, Severson, & Seeley, 2010). Universal screening of all students for behavioral and mental health needs are based on three approaches: teacher referral, proactive universal screening, and intervention-based identification (Walker et al., 2010).

Some critics have questioned the role of schools in providing mental and behavioral health services. However, the US Department of Education (2006) stated: "By all indicators, the need for mental health services has been more not less. Not everybody is in agreement that schools should be doing this. The long and short of it is there is some confusion about what constitutes mental health." The US Department of Education's stance on behavioral management and SMHS suggests a universal public health approach

to mental and behavioral health services, which has been reiterated in the President's New Freedom Commission Report that called on schools to provide more comprehensive behavioral RtI/MTSS and SMHS in order to decrease the prevalence of more serious mental health disorders. Further, behavioral RtI/MTSS and SMHS allows for collaborative relationships and coordinating of care between schools, families, and community agencies (Adelman & Taylor, 2004; Walker et al., 2010). The potential for powerful and effective behavioral services and SMHS is especially promising given that students spend more of their awake time at school than they do anywhere else. In summary, while some critics declare that schools are not appropriate places for behavioral management and SMHS and cite the lack of resources of schools as barriers to behavioral and SMH services, the fact of the matter is that schools are ideal places for behavioral and social-emotional interventions to take place and they are 'ipso de facto' the widest reaching and most impactful institution for service delivery.

Academic and Behavioral RtI/MTSS Improves Student Engagement. Given the federal mandate (IDEIA, 2004) to provide preventative and proactive strategies in education, the RtI/MTSS model holds promise in increasing student engagement in K-12 education. Educational research has shown that academic achievement is highly correlated to academic engagement and specific classroom factors such as time spent on academic subjects and opportunities for students to respond in learning environments (Greenwood, Delquardi, Stanley, Terry, & Hall, 1985, 1986; Soukup, Wehmeyer, Bashinski, & Bouvaird, 2007). An effective RtI/MTSS system that implements evidence-based academic, behavioral, and mental health strategies increases student engagement and conversely, student engagement improves academic, behavioral, and emotional

functioning (Watson, Gable, & Greenwood, 2011). When academic evidence-based interventions are implemented in an RtI/MTSS model, then student engagement increases because it ensures learning; put simply, academic achievement is incompatible with disengagement. Similarly, student engagement is incompatible with problem behaviors. Thus, an effective behavioral RtI/MTSS system promotes and increases student engagement by improving classroom environments and positive behavioral supports. However, there remains a need to better understand behavioral engagement in the classroom in order for an effective RtI/MTSS model to be implemented. As Watson et al. (2011) writes, “For RTI to be an efficient and effective problem-solving model that focuses on screening, early intervention, and prevention, data collected must be comprehensive and come from multiple sources that together provide information on student academic achievement, student behavior, teacher behavior, and instructional environments” (p. 32).

Student Engagement

Student Engagement and Academic Outcomes. Student engagement in the classroom has long been viewed as a beneficial construct that leads to desirable school outcomes (Carter, Reschly, Lovelace, Appleton, & Thompson, 2012). Over the last few decades in educational research one of the most consistent findings is the correlation between academic engaged time and learning (Winn, Menlove, & Zsiray, 1997). Fisher and Berliner (1985) found that student academic achievement is strongly correlated with academic learning time for students both with and without disabilities and these findings have been consistently repeated and replicated. For example, student engagement has been shown to be a good predictor of students’ academic achievement and school

completion (Connell, Spencer, & Aber, 1994; Furrer & Skinner, 2003; Skinner, Kindermann, & Furrer, 1998). Further, in a study by Klem and Connell (2004), students who self-reported higher levels of engagement were more likely to have increased attendance rates and higher grades. In contrast, higher levels of self-reported *disengagement* have been found to lead to more disruptive behaviors that impede learning, lower grades, reduce ambitious educational goals, and increase levels of high school drop out rates (Kaplan, Peck, & Kaplan, 1997). Despite the myriad ways student engagement is defined and measured, a review of literature indicates that student engagement consistently demonstrates positive academic outcomes across a wide variety of student outcome measures (e.g., higher standardized testing performance, higher grades, and increased student high school graduation rates) (Finn & Rock, 1997; Fredricks, Blumenfeld, & Paris, 2004; Sinclair, Christenson, Evelo, & Hurley, 1997).

Defining Student Engagement. A review of educational research suggests the construct of student engagement is nebulous. In fact, Appleton, Christenson, and Furlong (2008) suggest that a clearer conceptualization of the construct of student engagement needs to be developed through data supported research that validates the construct. Similarly, Blumenfeld (2006) stated that the construct of student engagement should be clarified among researchers since student engagement is an important and widely cited construct that leads to beneficial outcomes (Appleton et al., 2008). In a review of literature, much of the research on student engagement has been with college-age students. In fact, the National Survey of Student Engagement (NSSE) conducted one of the largest studies on student engagement. NSSE measures student engagement as a means to assess the quality of college experiences, which presumes that engagement

necessarily leads to the quality of the institution. The NSSE research suggested that there are seven components that contribute to student engagement in college: level of academic challenge, enriching educational experiences, student-faculty interaction, active and collaborative learning, and supportive campus environment (Kuh, 2003). The NSSE research aims to increase the environmental factors that contribute to engagement at the university level. Much less has been studied in early elementary classrooms in regard to student engagement and how it contributes to academic outcomes.

Cognitive scientists have also studied the construct of student engagement and have looked at how cognitive domains such as memory, language, and attention inform engagement in the classroom. In his book, *Why Don't Students Like School?* cognitive scientist Dan Willingham (2009) asserts that classroom teachers need to create certain “conditions” within the classroom to assist with student engagement. Willingham (2009) stated that students are naturally curious learners, but that curiosity is a “fragile” construct. His book outlines practices teachers should employ to ensure that students are engaged in learning; specifically, Willingham (2009) states that teachers should (1) Be sure that problems presented can be solved, (2) Respect students’ cognitive limits (e.g., limits on working memory), (3) Clarify the problems to be solved, (4) Reconsider when to puzzle students, (5) Accept and act on student variation in student preparation, (6) Continually change the pace of instruction, and (7) Keep a diary (i.e., teacher reflective practice) (Willingham, 2009, p. 15-17). Cognitive scientists have utilized knowledge of learning processes of the brain to suggest best teaching practices to increase student engagement. However, much of these practices are based on cognitive theories and have not been vetted in quantitative educational research. Similar to research conducted by the

NSSE, student engagement is often described in relation to teacher practice and producing engagement through environmental components and teaching strategies as opposed to defining the construct itself.

One of the first literature reviews conducted by psychologists on student engagement was in secondary schools by Mosher and MacGowan (1985) who noted the lack of research on student engagement in K-12 education and how student *disengagement* was more easily operationally defined (e.g., academic failure rate in grades, school drop out rate). Early research on student engagement identified key characteristics that conceptualize student engagement in secondary schools. Specifically, Mosher and MacGowan (1985) reported four identifiable “psychological characteristics” that are imperative components in the student engagement conceptual framework; they noted that student engagement is 1) the attitude that leads to participation in school activities, 2) has multiple complex “interactive determinants”, 3) has an impact on student academic outcomes and student behavior, and 4) should be conceptualized through a product of longitudinal research. Appleton et al. (2008) stated that while this 20-year old research was an important first step in conceptualizing student engagement, these characteristics lack clarity and more research is needed to operationalize and measure student engagement in schools. Appleton et al. (2008) further noted that student engagement is a multidimensional construct that almost always includes behavioral and psychological components and to a lesser degree includes cognitive and academic components.

Gettinger and Seibert (2002) described student engagement in a more formulaic fashion and stated that “academic learning time” (ALT) is defined as the “proportion of

instructional time allocated to a content area during which students are actively and productively engaged in learning” (p. 774). They further described the construct of student engagement into two categories: procedural engagement and substantive engagement (Gettinger & Seibert, 2002). *Procedural engagement* is described as being observable in nature. Gettinger and Seibert (2002) described procedural engagement as being more external where students are observed to engage in the behaviors of learning (e.g., participating in routines, completing tasks). *Substantive engagement* is more internal and describes the mental effort and active involvement in completing tasks (Nystrand & Gamoran, 1991). Nystrand and Gamoran (1991) stated that substantive engagement is the “sustained commitment to and engagement in the content of schooling” (p. 262). Fredricks et al. (2004) noted that substantive engagement is the student’s “investment” in the learning process and can be measured through student academic outcomes (e.g., words read per minute, reading comprehension, math performance). In other words, procedural engagement is the *going through the motions* of schooling (e.g., showing up, looking at the teacher, completing worksheets), but substantive engagement is less observable and includes mental effort factors, such as persistence, which is more difficult to measure in a quantifiable and observable fashion.

While conceptualizations of student engagement vary among educational researchers and theorists, there are three main types of engagement that have emerged in more recent research on student engagement: *emotional*, *cognitive*, and *behavioral* (Betts, Appleton, Reschly, Christenson, & Huebner, 2012; Finn & Zimmer, 2012; Fredricks et al., 2004; Reschly & Christenson, 2012). Fredricks et al. (2004) described *emotional* engagement as more of an internal process where a student’s engagement is measured by

the types and amount of positive and negative feelings they experience. *Cognitive* engagement includes the level of investment a student has in learning, which then affects effort (Fredricks et al., 2004). Finally, *behavioral* engagement includes observable behaviors in the classroom such as paying attention, participation in classroom activities, and time on-task (Fredricks et al., 2004).

These conceptualizations of student engagement suggest that *emotional* and *cognitive* engagement are more internal processes that are likely to be best measured through self-reports and interviewing techniques. In contrast, *behavioral* engagement is suitable for quantifiable, objective measurement because it includes overt behaviors. In fact, Walker and Steverson (1994) defined academic engaged time into three observable categories: 1. *Attending* to the instructional material and task, 2. *Motor Responses* that are appropriate to the task at hand (e.g., writing, looking at the teacher, reading), and 3. *Asking for assistance* in an appropriate manner. These academically engaged behaviors include eye contact with the teacher, reading a textbook, and engaging with the instructional material through problem solving, discussion, and social engagement with peers around active participation in classroom learning. Similarly, Kauchak and Eggen (1993) defined academic engaged time as being synonymous with “time on-task.”

Student Engagement and Construct Validation. Prior researchers have suggested that internal or substantive engagement is difficult to measure because internal engagement requires more inference from an outside observer and often relies solely on self-report measures (Pintrich, Wolters, & Baxter, 2000; Winne & Perry, 2000). A study conducted by Spanjers, Burns, and Wagner (2008) examined the relationships between direct behavioral observations of student engagement, academic achievement, and

student self-reports of engagement. Their research questioned how observable behavioral engagement related to internal or substantive engagement. Specifically, Spanjers et al. (2008) studied the generalizability of observable behavioral engagement and the relationships between self-reported engagement and academic performance. In their study, Spanjers et al. (2008) conclude that more research is needed to validate behavioral engagement and its relationship to academic performance because their findings found a low correlation between behavioral observations of time on-task and academic performance. These findings are in opposition to the large body of research that asserts that student behavioral engagement is highly correlated with student academic achievement. However, one of the interesting questions they posed was: Does observable engagement (time on-task) correlate with self-report measures of engagement? Their findings indicated that small partial correlations are found for academic outcomes and behavioral engagement and self-reported engagement (Spanjers et al., 2008). These findings beg the question: How important is internal or substantive engagement versus behavioral engagement in producing positive academic outcomes? One of the limitations of the Spanjers et al. (2008) study, which was addressed in their article, is that they measured academic performance solely through reading comprehension passages where percentage of correct answers were calculated to assess academic performance. Their study did not consider more robust measurements of student academic performance, such as reading fluency (i.e., words read correctly per minute) or math calculation skills (i.e., digits correct per minute), which is considered to be a more valid general outcome measure that better predicts academic achievement (Hosp, Hosp, & Howell, 2007).

Measuring Student Engagement. Fredricks et al. (2004) conducted a meta-analysis on the number and types of measurements used to measure student engagement in upper elementary and high school students. Their findings indicated that there are 21 measures of student engagement with the majority of these measures being student self-reports. In fact, 14 of the 21 measures were student self-reports, 3 of the 21 were teacher reports, and 4 of the 21 were observational measures. Of the 4 observational measures, only 2 were systematic, standardized direct observations (SDOs) of students that used peer comparison as a relative measure of student engagement and the SDO measures were designed to observe an individual student, not overall classroom behavior (Fredricks et al., 2004). Thus, while current observational methods yield rich information on an individual's academic engaged time, there remains a need to establish reliable, normative benchmarks for academic engaged time on a larger scale to better interpret individual student data and develop a deeper understanding of classroom- and school-level trends in behavioral engagement. These benchmarks could provide a basis for interpreting data and evaluations of programs, supports, or initiatives. Further, establishing normative benchmarks for academic engaged time would allow one to quickly and easily understand how individual schools, classrooms, or students are performing relative to same-grade peers within their locale. Normative benchmarks of academic engaged time, as measured by on-task, off-task, and disruptive behaviors, could also serve as a tool to help identify classrooms in need of additional support, facilitate educational planning (e.g., times of day best for instruction), and could also be used in research settings (Gresham, 1981; Haynes & Wilson, 1979; Klein, 1979).

Simple View of Learning and Academic Engagement. Educators, scientists, and philosophers have long debated and attempted to define learning and the learning process. While student learning is a complex process that can be described by a teacher as the moment of comprehension (an “aha” moment) or by a neuropsychologist’s explanation of growing dendrites in the brain, learning in schools can be concretized as a product of the amount of time devoted to instruction multiplied by the amount of time students are academically engaged in the learning (Rathvon, 2008). Thus, learning can be represented as the following equation: Learning (L)= Time Devoted to Instruction (TDI) x Academically Engaged Time (AET). This equation for learning as depicted in Figure 1 can be used as a quantifiable way to measure the conditions necessary for an effective learning environment. Although $L=TDI \times AET$ may represent an overly simplified definition of learning, given the vast array of definitions and complexities involved in learning, education fundamentally involves teachers teaching and students attending school and being engaged in the instruction. In addition, as previously discussed, it is well established that the amount of time students are academically engaged in the classroom is intricately linked to student academic performance (Baugous & Bendery, 2000; Prater 1992). Further, the construct of student engagement is nebulous, at best, and often defined through teacher and school practices that increase engagement, rather than defining the construct itself. While further research and study should be given to validating the construct of student engagement, for this study, student engagement is defined as AET and is based on the Rathvon’s $L=TDI \times AET$ formula. Therefore, given this simplified view of learning, educators must focus on maximizing the amount of instruction delivered throughout the school day and ensure that students are engaged

when instruction is delivered. Finally, given the strong relationships between student engagement, academic engaged time, and on-task behavior, these terms will be used synonymously for the purpose of this study given that academic engaged time was measured by observable and quantifiable on-task behaviors.

Systematic Direct Observations (SDO)

Observations are a commonly utilized assessment tool used to evaluate student behavior. Within schools, there are two major types of observational assessment: 1) Naturalistic/descriptive observations where there are no predetermined specified behaviors being assessed, which often include anecdotal narratives, and 2) Systematic direct observations (SDOs) where the observer has a predetermined set of behaviors that are observed in a systematic and explicit manner (Salvia & Ysseldyke, 2004). In a survey of over 1,000 school psychologists, SDOs were one of the most frequently used observational assessment tools (Wilson & Reschly, 1996). Shapiro and Heick (2004) surveyed school psychologists ($n=648$) and found that 69% of the respondents used SDOs in about 4 out of 10 of their cases that were referred for emotional or behavioral problems. In contrast, Volpe and McConaughy (2005) noted that anecdotal or naturalistic observations were more widely used when compared to structured observations. Volpe and McConaughy (2005) stated that there are several flaws when exclusively relying on anecdotal observations; they write that naturalistic observations lack precise definitions of target behaviors, provide only descriptive information instead of quantitative data that can be analyzed, cannot be tested for reliability (i.e., interobserver agreement), and there is no way to test the validity of the observations (p. 451). Volpe and McConaughy (2005) point out that SDOs measure specific target behaviors, include behaviors that are

operationally defined, have standardized coding procedures, yield quantitative scores, and can be tested for reliability and validity across observers, settings, and time. Similarly, according to Salvia and Ysseldyke (2004), major characteristics of SDOs are that they measure specific behaviors that are predetermined and operationally defined; SDOs use standardized procedures that are conducted at specific times and in specific environments and SDO assessments are standardized in a scoring system that allows for data to be analyzed quantitatively.

Many of the commonly used standardized SDO instruments look at individual students using well-defined operational categories of behavior. Common behavioral categories in SDO data are on-task, off-task, and disruptive behaviors (Merrell, 2010). In order to compare an individual's classroom behavior, educators most often use what Sattler & Hoge (2006) refer to as "informal norms." For example, Alessi (1980) suggested that an individual's classroom behavior could be compared to a peer of the same sex who is considered more typical of the class make-up, which serves as a "micro-norm." Using local, classroom norms for peer comparison serve as a local normative benchmark to compare a student's classroom behavior to their peers within the same classroom. Using local or micro-norms thus allows an observer to create their own benchmarks within a classroom by calculating a discrepancy ratio (Sattler & Hoge, 2006). Volpe and McConaughy (2005) noted that SDOs can be used in two ways: nomothetic or idiographic assessments. In a nomothetic assessment, an individual's functioning is compared with other groups, and in idiographic assessment, an individual's functioning is examined to inform interventions (Volpe & McConaughy, 2005). According to Shapiro (2004), SDO data has often been used in functional behavioral

assessments that examine an individual's behavioral functioning to drive hypotheses and intervention planning (idiographic assessment).

While SDO instruments provide peer comparison methods or idiographic assessments as described above, in a review of previous educational research, there appears to be little research on how SDOs should be used to interpret student data, especially in nomothetic assessments. In fact, Volpe and McConaughy (2005) write, "Although there is now evidence that school psychologists use systematic direct observations, it is not at all clear what criteria school psychologists use to select particular observational procedures, whether they use and interpret the system correctly, nor whether they are aware of observational systems available to them. Furthermore, there are relatively few published studies concerning the psychometric properties of commercially available direct observation systems" (p.452). Similarly, Briesch et al. (2010) noted that psychometric evidence for behavioral assessment methods are far less developed when compared to academic and ability assessments and this is an area that especially needs to be developed given the importance of making data-based decision in a RtI/MTSS system delivery model. Hintze (2005) also noted that psychometric properties of behavioral assessments are far less developed when compared to other instruments. In their study, Briesch et al. (2010) compared SDOs and behavior ratings scales and found both were reliable and sensitive tools to assess student behavior; their study did find variance in behavioral measurements among raters and SDO data, which was affected by changes in student behavior across settings and time.

Hintze (2005) writes that behavioral assessment is necessary in a RtI/MTSS delivery service model, yet most behavioral assessments are not designed to be frequently

administered (e.g., most tend to be lengthy rating scales) and the available behavioral assessments often do not take into account the time and resources needed to feasibly collect data. Hintze (2005) writes, “A pressing need therefore exists to identify [behavioral] measures that are both technically adequate and feasible for applied use in problem-solving assessment” (p. 409). Hintze (2005) proposes two behavioral assessments that could feasibly be used in RtI/MTSS that would allow for frequent and reliable measurement of student classroom behavior: SDOs and direct behavior ratings (DBRs). Hintze (2005) notes limitations in regards to SDO data; specifically, Hintze (2005) asserts that SDOs often require unavailable resources (i.e., school staff that are trained to conduct observations), and they provide only a small snapshot of behavior at a specific time and setting. More research is needed to study the feasibility of conducting SDOs as part of the data-based decision making process in RtI/MTSS and more research on the variance of SDO data within students, classrooms, and schools is needed to further establish the psychometric properties of SDO data and its reliability in interpreting student classroom behavior.

Importance of Establishing Behavioral Benchmarks of Classroom Behavior

Early elementary grades are critical years for students to acquire core foundational skills that are needed for future success and schools play a pivotal role in preventing academic and behavioral difficulties (Kame’enui & Simmons, 1998; Walker, 2000). Since educators play an important role in preventing academic and behavioral difficulties, they must shift the focus from the traditional ‘wait-to-fail’ model where they intervene at the individual level in a reactive system to focusing on *all* students using a population-based proactive approach (Shapiro, 2000). In order to effectively implement a

population-based RtI/MTSS approach, educators must have adequate and reliable normative data in order to best interpret and measure student performance using universal screening and progress monitoring instruments. Early detection of academic and behavioral difficulties is imperative because as academic and behavioral problems increase, students are less responsive to interventions since the gap between their performance and grade-level expectations will grow wider and become more difficult to close (Kratochwill, 2007). Thus, normative data plays a key role in interpreting and measuring student performance, so that educators have reliable and valid tools in order to effectively implement an RtI/MTSS model.

Academic normative data and academic universal screeners and progress monitoring probes are well established within early elementary classrooms; curriculum-based measurements (CBMs) of reading, math, and writing skills have been widely researched and are frequently used within early elementary classrooms (Hosp et al., 2007; Fuchs, 2004). Educators have access to reliable curriculum-based measures that are based on normative data that allow them to effectively screen and monitor students, which is essential in an RtI/MTSS data-based decision making model. Academic normative data not only contributes to understanding and interpreting individual student performance, it also provides valuable information about classroom, school, and district level performances. As previously described, universal screening data allows educators to better understand the quality of their Tier 1 supports. Further, having normative data allows educators to have a reference point to gauge Tier 1 instruction and evaluate Tier 1 improvements. Academic curriculum-based measurements are often administered to calculate classroom, grade, and school wide averages of academic performance as a way

to inform instructional practices (Cook, Volpe, & Livanis, 2010). For example, if universal screening data indicated that 50% of students in a particular classroom were reading below grade level according to a curriculum-based measurement of reading, then this would indicate that Tier 1 instruction needs to be adjusted.

Active monitoring of Tier 1 instruction has become an integral component in evaluating the effectiveness of instruction (Kupzyk, Daly, Ihlo, & Young, 2012). Further, teacher evaluation systems are increasingly using “value-added” models (Harris, Ingle, & Rutledge, 2014). For example, the new federal program, *Race to the Top*, rewards teachers and administrators for increased student achievement, which has been described using the term “value-added” (Harris et al., 2014). Another example of increased accountability using performance data is in Florida where state legislation asserts that teacher promotion, tenure, dismissal, and compensation should be based on quantitative “value-added” results (Harris et al. 2014). Teacher evaluation systems not only use “value-added” academic achievement data to measure performance, but student engagement is often cited as an evaluative component in value-added teacher evaluation systems (Danielson, 1996). For example, the Danielson Framework, which is a popular teacher evaluation system often used in schools, outlines criterion in which teachers are measured, including planning and preparation, the classroom environment, instruction, and professional responsibilities (Benedict, Thomas, Kimerling, & Leko, 2013). More specifically, the Danielson Framework criterion #1, “Centering instruction on high expectations for student achievement”, explicitly states that engaging student in learning is a core competency for effective teaching (Danielson, 1996).

While the link between engagement and academic performance is widely accepted among educators, previous research has not established a large body of quantitative data for measuring academic engaged time, which could be assessed through measuring on-task, off-task, and disruptive behaviors in early elementary classrooms. On-task behavior is incompatible with off-task behavior; meaning it is impossible for a student to be off-task when they are on-task (Shapiro, 2011). Disruptive behaviors are also a measure of behavioral engagement because disruptive behaviors within the classroom impact teachers' ability to deliver instruction and students' opportunities to maintain on-task behaviors. Further, while psychometric properties of more summative cognitive and academic achievement assessments are well researched, there is far less research on behavioral assessment (Briesch et al., 2010). The lack of research on the monitoring of student behavior is surprising given that IDEIA (2004) requires technically adequate tools to gather information about student behavior (Briesch et al., 2010; Fuchs & Fuchs, 2006). Schools are expected to serve students with both academic and behavioral difficulties, yet most universal screeners and progress monitoring tools for teacher and student performance are academic based, not behavioral (Cook et al., 2010; Lane, Oakes, & Menzies, 2005). In fact, while there have been huge strides in the development and implementation of academic progress monitoring of students such as the use of curriculum-based measurements, much less is known about observational behavioral screeners. Because of the major focus on academic progress monitoring and the lack of behavioral progress monitoring tools, it is proposed that there remains a need to better evaluate and interpret overall classroom behavior. For example, Hintze (2005)

states that there is a need for behavioral assessment to “demonstrate reliability, sensitivity, and feasibility in the solving problem process” (p.259).

Purpose of this Dissertation Study

An additional lens for understanding classroom behavioral trends should be established. The primary purpose of this study was to establish normative benchmarks of early elementary students’ academic engaged time and disruptive behavior to be used for a variety of practical purposes in the schools with regard to establishing school- and class- wide goals, monitoring individual student progress in response to intervention, and evaluating educators as part of a larger portfolio of assessments. These normative data were examined to determine the extent to which student classroom behavior varies across contexts (e.g., classrooms, schools, and districts), and to examine school-level variables associated with behavioral engagement. Such descriptive analyses of the participating students ($n=6,592$) will contribute to the research on behavioral data of students in early elementary classrooms and will assist researchers to better interpret student behavior and engagement in the classroom, which is an essential component in a RtI/MTSS educational service delivery model. A secondary aim of this study was to conduct school-level analyses ($n=56$) to examine factors at the school-level that impact academic engaged time and disruptive behavior. This school-level analyses will assist researchers and educators in identifying and targeting that may need to be the focus of prevention and intervention efforts, which could influence the allocation of resources and educational policy. Thus, this dissertation study aimed to answer the following questions:

1. Research Question #1: What is the Kindergarten, First, and Second grade normative data for Academic Engaged Time (AET) and Disruptive Behavior (DB)?
2. Research Question #2: What is the variability in normative data across classrooms, schools, and districts?
3. Research Question #3: To what extent are school-level variables (percent of students qualifying for Free and Reduced Lunch [FRL], percent of white/non-Hispanic students, and academic achievement data) associated with school-level norms regarding Academic Engaged Time (AET) and Disruptive Behavior (DB)?

With regard to the first research question, it is hypothesized that the mean academic engaged time and disruptive behavior is fairly consistent across grade levels with students demonstrating more on-task behaviors and less disruptive behaviors. With regard to the second research question, it is predicted that there will be more variability on measures of dispersion across classrooms indicating that there is more variability within schools than between schools. Finally, with regards to the final research question, it is predicted the school-level variables will be significantly associated with observed academic engaged time and disruptive behaviors. Specifically, it is hypothesized that academic achievement, percent of students qualifying for FRL, and percentage of white/non-Hispanic students will predict academic engagement and disruptive behaviors. These hypotheses are supported by prior research that has shown that socioeconomic status is linked to increased social-emotional and behavioral needs among students, resulting in less engagement and more disruptive behavior (Maag & Katsiyannis, 2010). Non-white status is also likely to relate to academic engagement because prior research

has indicated that there is an academic achievement gap between white and minority students (US Department of Education, 2000). Sirin (2005) cited factors that contribute to the inequalities between white and minority students; minorities are more likely to live in low-income households, have a higher number of external stressors, have lower parental education, and attend underfunded schools. While research connecting socioeconomic factors and minority status with academic achievement has been well established in educational research, there is little research on how school-level socioeconomic factors and minority status relate to academic engagement.

Chapter 3: Methods

Setting and Participants

Participants were Kindergarten through Second graders in participating school districts in the Puget Sound and Mesa, Arizona areas. School districts were recruited through the Second Step Study in the Spring of 2012. The Second Step Study attained approval from the University of Washington's Institutional Review Board and consent was also received from the participating school districts, teachers, students, and parents. Data from the 61 schools in the North Kitsap, Renton, Federal Way, Mukilteo, Lake Washington, and Mesa School Districts participated in this study. The author was personally responsible for collecting data for eight of the school participating in the study, and she was intimately involved in recruitment, IRB approval, and organization of the study. Of the 7,419 students in the study, observational data was collected on 6,592 of the participants due to withdrawal from the study, students moving out of the area, or absences on the date the observational data was collected, resulting in a 11% attrition rate. Of these participants, 3,135 were Kindergartners (42.3%), 3,854 were First graders (51.9%), and 430 were Second Graders (5.8%). With regard to socioeconomic status, approximately 61% of the students were receiving free and reduced lunch. The ethnicity breakdown of the students was as follows: Caucasian/White – 36.3% ($n=2,694$), Black or African American – 6% ($n=447$), Asian-American- 9.5% ($n=702$), Native Hawaiian or other Asian/Pacific Islander – 9.5% ($n=72$), American Indian or Alaska Native- 2.8% ($n=210$), Hispanic – 22.5% ($n=1,669$), More than one race- 5.1% ($n=379$), and Unknown- 16.8% ($n=1,246$).

Design

This study represents part of a larger collaborative effort between the Committee for Children, University of Washington, and Arizona State University to evaluate the impact of two years of implementation of the newly revised Second Step program in early elementary classrooms. This study utilizes data from the Second Step Study. Sixty-one elementary schools from two sites (UW – 41 and ASU – 20) were randomly assigned within their district to either the early start ($n = 31$) or delayed start ($n = 30$) conditions. Data for the larger Second Step Study was collected across six waves (e.g., baseline-Fall year 1, post1-Winter year 1, post2-Spring year 1, post3-Fall year 2, post4-Winter year 2, and post5-Spring year 2) on a variety of measures assessing student outcomes (e.g., behavior ratings, direct observations, academic performance) and classroom environment. For the purposes of this study, data from only the first wave in Fall 2012 will be analyzed. The data from Fall 2012 was the baseline data that occurred before the implementation of the Second Step curriculum. Data utilized in this study from Fall 2012 includes demographic information, classroom, school, and district placements, observational data of on-task, off-task, and disruptive behaviors, and reading and math achievement measures.

Measures and Procedures

Systematic Direct Observations. To record class-wide and individual student behavior, a behavioral observation system was developed based on the Behavioral Observation of Students in Schools (BOSS; Shapiro & Kratochwill, 2000). A Classroom Behavior Observation Form (CBOF) was developed based on the BOSS and was designed to record data on all students in a given classroom to produce both class-wide and individual estimates of behavior. An example of the behavioral observation form

used for this study is displayed in Figure 2. The three behavioral coding categories consisted of on-task behavior or academic engaged time (AET), off-task behavior (OFT), and disruptive behavior (DB). There was some overlap between the behavioral codes, as the coding CBOF procedure made it possible for a student to be coded as off-task/non-disruptive or off-task/disruptive since OFT was momentarily time sampled and DB was partial interval time sampled. Time sampling is further explained below. Coding steps for the CBOF were as follows:

1. Begin with a student at the front or back of the room and write student descriptor.
2. Complete six 10-second intervals on the student before moving to the next one.
3. Move clockwise to the next student and write student descriptor (30 seconds to complete).
4. Begin recording six 10-second intervals.
5. Repeat process for next student.

The coding system for the CBOF was provided in an Excel format and was inputted directly onto a laptop. If a laptop was not available, the observer used a clipboard and a printout of the CBOF form. In addition, a stopwatch was needed to keep track of time intervals. Behavioral definitions for the variables coded in the observation were as follows:

- ***Academic Engagement Time (AET) momentary-time sampling*** is defined as times when the student is working on the academic task at-hand or paying attention to the lecture by having eyes focused on the teacher. Examples of academic engagement included writing, reading aloud, raising a hand and waiting patiently, talking to the teacher or other students about assigned material, and

looking things up that are relevant to the assignment. **AET** was recorded at the very beginning of each interval, not throughout or during the middle of an interval.

- ***Off-Task (OFT) momentary-time sampling*** is defined as when a student is not engaged in the academic task at hand, because he or she is staring off, talking to other students about non-academic things, and being disruptive. **OFT** was recorded at the very beginning of each interval, not during the middle of it. Also, a student cannot be academically engaged and off-task at the same time; the two categories are mutually exclusive.
- ***Disruptive Behavior (DB) partial interval*** is defined as behaviors that are not related to the task at hand and are disruptive to learning or the classroom environment, but do not pose immediate danger to the other peers, teachers, or property (e.g., call outs, talking to peer when not permitted, out of seat, behavior that draws other peers off-task, playing with object).

Observations were conducted in 318 classrooms across the University of Washington and Arizona State University sites. Trained graduate students completed the observations during core academic instruction. Core academic instruction was defined as Reading, Writing, Math, Social Studies, or Science instruction. All of the observations were conducted in September and October of 2012; October 15th was the last day of data collection. Each student participating in the study was observed for two minutes total, divided into 10-second intervals. To obtain class-wide estimates of AET and DB, observers were instructed to begin with an identified student in the front or back of the classroom and systematically move to the next student to the left after each interval. After

the observers made their way through all students in the class, they repeated the same process until the observation time elapsed. A minimum of 12 intervals of data per student and roughly 300 total intervals across all students were obtained. Observations were between approximately 60-90 minutes. These observations allowed for the calculation of class-wide and individual student estimates of AET, OFT, and DB.

Prior to conducting the observations, graduate students were trained on the observation system. Before beginning baseline data collection, each student was required to reach at least 90% agreement during practice trials with an identified observer who served as the anchor measure. These preliminary observations were conducted in an elementary school not participating in the study and the data collected was intended solely to train graduate students on the observation measure. For the observational data included in this present study, inter-observer agreement (IOA) data consisting of two observers conducting the observation at the same time on the same students were collected on roughly 20% of the observation sessions. IOA was calculated using the point-by-point method, which consists of calculating agreement for each and every interval. This method has been shown to be an accurate estimate of the agreement between raters for direct observation systems with interval recording formats (Shapiro & Kratochwill, 2000). The results revealed that IOA averaged 88% (minimum = 72% and maximum = 100%), which was associated with a Kappa value of .71 and is considered to be an acceptable level of inter-rater reliability (Bailey & Burch, 2002; Viera & Garrett, 2005).

Curriculum-Based Measurements (CBMs). To assess academic performance, the following CBMs probes from the Aimsweb© program were administered: (a) oral

reading fluency (words read correct per minute) and (b) math calculation (number of digits correct in a minute and percent correct). Reading-CBM or oral reading fluency probes represent a standardized, general outcome measure of reading performance that is highly sensitive to students' response to instruction (Fuchs et al., 1999). Math-CBM or math computation probes have been shown to be a reliable and valid general outcome measure of overall mathematics computation (Thurber, Shinn, & Smolkowski, 2002). Students received grade-equivalent probes for both of these measures. Trained graduate students administered CBMs in all 318 classrooms.

The Reading CBM (R-CBM) is a one-minute timed reading of a short passage, which measures oral reading fluency (ORF). The Reading CBM was administered one-to-one with the student. The trained graduate students conducting the R-CBM used a stopwatch, a clipboard, a wipe cloth, the provided grade-leveled passage in a plastic sleeve, a dry erase pen, and their laptop or a recording sheet to record the student's overall words read correctly per minute (wcpm). The following are the directions for how the Reading-CBM was administered, which was adapted from Aimsweb©:

1. Place the Passage in front of the student.
2. Say, □ *“When I say ‘Begin,’ start reading aloud at the top of this page. Read across the page (point). Try to read each word. If you come to a word you don’t know, I’ll tell it to you. Be sure to do your best reading. Are there any questions?”* (Directions may be shortened.)
3. Start timing the student after you say “Begin.”
4. Follow along on your copy. Put a slash (/) through words read incorrectly.

5. At the end of one-minute, place a bracket (]) after the last word and say, “Stop.”
6. Thank the student for their effort.
7. Score and record the student’s words read correctly (WRC).

At times when the student read aloud, it could be difficult to decipher what counts as an error. In order to ensure the most accurate measure of their ability when students hesitated or ‘got stuck’ or other errors occurred, the following standardized assessment procedures were used to ensure accuracy:

- If a student pauses or struggles for 3-seconds, give them “the word” and mark it as incorrect.
- Self-corrections are marked as correct if the student self-corrects within 3-seconds of reading the word.
- Mispronounced words or similar-substituted words are marked as incorrect.
- Omitted words are marked as incorrect.
- Repetitions, dialectical differences, or additional words are neither scored correct or incorrect, so ignore them.
- Do not correct the student’s reading errors.
- Mark where the student ends in one-minute, but use your best judgment when it is polite to let the student finish before saying “Stop.”
- If a student speed reads (i.e., very fast and without expression), tell them that this is not a speed-reading test and to begin again.
- If there is a disruption during the assessment, re-administer the Reading CBM with a different grade-leveled passage.

Of note, when R-CBM's were administered in Kindergarten classrooms, many of the students were non-readers or non-fluent readers. Trained graduate students were instructed to record the number of words that they can read correctly within a minute, which was often a sight word search within the passage. Kindergartners needed encouragement at times to try and read any word on the page. Graduate students were trained to use their best judgment; if the child told the graduate student that they could not read or could not read any words on the page, their score was recorded as zero.

The Math-CBM assessed math computational skills (M-COMP). It was an eight-minute test that was group administered. Graduate students used a timer, sharpened pencils, and enough copies of the M-COMP testing sheet for the entire class.

Instructions for the Math-CBM were adapted from Aimsweb© and were as follows:

1. Say, *“We are going to take an 8-minute math test. Read the problems carefully and work each problem in the order presented, starting at the first problem on the page and working across the page from left to right. Do not skip around. If you do not understand how to do a problem, mark it with an X and move on. Once you have tried all the problems in order, you may go back to the beginning of the worksheet and try to complete the problems you marked. Although you may show your work and use scratch paper if that is helpful for you in working the problems, you may not use calculators or any other aids. Keep working until you have completed all of the problems or I tell you to stop. Do you have any questions?”*
2. Answer any questions the students may have.
3. Pass out tests.

4. Say, “Begin.”
5. After 8 minutes, say, “Stop and put your pencils down.”
6. Collect sheets and thank the class.

If a student asked a question(s) or requested help during the testing, the graduate student was instructed to tell them to read the directions again and work the problems as best as they can. If they saw a student skipping around the worksheet and not completing the math problems in order, the graduate student was instructed to coach the student to try and work each problem in order.

School-level Data. School-level data were gathered from publically available on-line sources (e.g., NCES website, school district websites) and directly from the school districts. School-level data collected included the number of students in the school and classrooms, racial/ethnic composition of students, percentage of students receiving free or reduced-price lunch (FRL), percentage of students who were English Language Learners (ELL), percentage of students in special education, and percentage of students meeting statewide reading standards.

Data Analytic Strategies and Research Questions

Descriptive statistics were calculated as the primary means to analyze the data in order to describe the normative characteristics of the data (Cohen, Manion, & Morrison, 2003). Measures of variability were calculated by determining the standard deviation amongst scores in order to give a sense of the average dispersion of scores within a data set. To account for inevitable sampling errors, standard errors of mean and confidence intervals were calculated. These statistics were used to examine individual, classroom, school, and district differences in SDO data by examining the variability in on-task, off-

task, and disruptive behaviors at each level. Stepwise linear and logistic regressions analyses were employed to analyze the degree to which school-level variables were associated with AET and DB. Specifically, the data analytical approach for each research question is outlined as follows:

Research Question #1: What is the Kindergarten, First, and Second grade normative data for Academic Engaged Time (AET) and Disruptive Behavior (DB)?

For this question, the approach utilized by Agnoff (1984) was used to establish normative benchmarks. This entailed assessing normality, establishing mean, median, mode, and range as well as standard deviations and percentile ranks of the total sample and grade levels. Agnoff (1984) states that the term normative has two meanings: 1) the actual performance of a well-defined group of individuals, or 2) the standards or goals of performance. For the purpose of this study, normative data refers to the actual performance of students' behavioral functioning as measured by SDOs. Norms for this study were developed through determining an individual's relative standing within a defined population (Agnoff, 1989, p. 39). Agnoff (1989), Conrad (1950), and Schrader (1960) outlined general conditions needed to construct norms that are summarized briefly below:

1. The characteristic measured must be ordinal (e.g., able to be organized from low to high).
2. The characteristic being studied should be operationally defined.
3. The measure being used should examine the same characteristic in all of its scores.

4. The group being examined should be appropriate to the measure being used; in other words, the measure itself should be appropriate for the population being studied.
5. Data should be provided that breaks down distinct norms within populations.

In order to construct norms, specific descriptive metrics were computed including measures of central tendency (means, median, and mode) and variability (standard deviation, range, percentiles). This data helped establish benchmarks that facilitated in the interpretation of individual and aggregated (class or school) student data. Finally, correlation matrixes for student and school-level variables were constructed to further examine the relationships between behavioral engagement, academic achievement, and demographics.

Research Question #2: What is the variability in normative data across classrooms, schools, and districts?

For the second research question, variability in benchmark data within classrooms, schools, and districts were plotted graphically to demonstrate the distributions. Thus, histogram plot graphs were constructed to visually depict the variability of means within regions and districts. Further, examples of school histograms will visually represent the variability of means of classrooms within schools. District and school means with 95% with corresponding confidence interval bars were also graphed for AET and DB. The 95% CI bands were compared visually and when the CI bands did not overlap between specific levels of analysis (i.e., districts, schools, or individuals), the conclusion was that there was significant variability within that level of analysis.

Moreover, the graphs were visually analyzed to examine the variability of engagement and disruptive behaviors for all the participating districts and schools in the study.

Research Question #3: To what extent are school-level variables (percent of students qualifying for Free and Reduced Lunch [FRL], percent of white/non-Hispanic students, and academic achievement data) associated with school-level norms regarding Academic Engaged Time (AET) and Disruptive Behavior (DB)?

For research question 3, multiple linear and logistic regression models were performed to determine which school-level variables (FRL status, percentage of white/non-Hispanic students, and academic achievement data) were meaningfully associated with AET and DB. As depicted in Figures 14 and 15, the AET dependent variable is negatively skewed with most schools having over 80% of their students' time actively engaged, and the DB dependent variable is positively skewed with most schools having less than 10% of their students' behavior coded as disruptive. In order to meet the assumption of normality, a skewness score must not be above 1.0 or below -1.0. Because school-wide AET and DB did not meet normality criteria, transformations of the data were attempted in order to normally distribute the data. For example, the author attempted to meet the assumption of normality through log, square, cubic, and quadratic transformations. As Figures 14 and 15 demonstrate, the AET and DB outcomes were measured 0-1 (0-100% engagement or disruptive behavior) and the data was 'stacked' positively or negatively; it was not possible to meaningfully exclude any of the data since there were no extreme outliers. None of the initial statistical approaches, such as transformations, trimming means, or truncating the data, resulted in a normal distribution.

Therefore, statistical methods that would require normality would not be valid for this analysis.

The school level data in this study were nested within districts and regions and school-level variables were inter-correlated. Given that the outcome variables were not normally distributed and to account for school-level nesting and clustering within districts, a generalized estimating equation (GEE) method (Zeger & Liang, 1986) was used to run the linear regression models. GEE regression was used because GEE does not require normally distributed outcome data (Zeger & Liang, 1986). A Poisson distribution was considered, but not utilized because the AET and DB outcome variables were not frequency/rate data, and the AET and DB outcome variables were standardized into an overall percentage to make all school AET and DB on the same scale. A linear regression fit using GEE modeling to relax the normality assumption was utilized for each outcome variable. Fitting the regressions using the GEE method with robust standard error specifically addressed the concern of non-normality of the outcome variables since this approach relaxes this assumption and is valid for this study's outcome types (Zeger & Liang, 1986). To approximate fitting standard multi-level models which include random effects for districts, the GEE models were fit using an exchangeable correlation structure on districts to account for district nesting. Further, GEE accounts for clustering, or intercorrelated data (Hardin and Hilbe, 2002). Hardin and Hilbe (2002) noted that using a GEE approach was first developed in the mid-1908s to address longitudinal and cluster data instead of using a traditional General Linear Model (GLM). Liang and Zeger (1986) developed the GEE approach to address dependence in data. Ziegler (2011) stated that the GEE approach has become increasingly popular in recent years because it overcomes the

“classic” assumptions of normality and independence. While GEE statistical methods are most often used in the biostatistics field, this statistical method approach remains a valid approach in other fields. Thus, linear regressions fit using GEE with exchangeable correlation structure were employed to account for nesting/clustering, and non-normality of the AET and DB outcomes. This analysis was conducted using the R statistical software. The author ran separate linear regression models for each school-level variable and each outcome with adjustment for regional effects (University of Washington versus Arizona State University sites). However, given the non-normal distributions of AET and DB variables and the stacking of outcomes towards high percent AET or low percent DB as depicted in Figures 14 and 15, the author ran secondary analyses using logistic regression by dichotomizing the variables into high and low and conducting the analysis on a binary outcome variable (high or low engagement; low or high disruptive behavior). The logistic regression models were fit using GEE with exchangeable correlation structure to handle the nesting of the outcomes similar to the continuous outcome analyses.

Descriptive statistics were examined to determine meaningful binary categories of AET and DB for the logistic regressions. Since there is no prior research on quantitative behavioral engagement as measured by SDO data, the author used a cut-off of 85% to differentiate between schools with higher academic engagement and schools with lower academic engagement and the cut-off of 10% for disruptive behavior after examining the descriptive statistics. Further, for the sake of meaningful interpretation, FRL status was categorized into Title I and non-Title I schools. In the United States, schools are considered Title I schools if 80% or higher of the school population meets criteria for

Free and Reduced Lunch. Title I status is an indicator of socio-economic demographics of the school and is a common metric used in educational research (US Department of Education, 1991; NCLB, 2001). To meaningfully interpret school race and ethnicity demographics, a quartile analysis of school-level white/non-Hispanic percentages was conducted and resulted in the following categories: $<30\%$, $30-<55\%$, and $\geq 55\%$. The reading and math independent variables were based on curriculum-based measurements that measured words correct per minute (wcpm) for reading achievement and digits correct per minute (dcpm) for math achievement. Finally, it should be noted that there were missing demographic data for 5 of the 61 schools participating in the study. While descriptive statistics examined in the prior research questions were able to examine all 61 schools, complete demographic data were obtained for only 56 schools to address the final research question.

Chapter 4: Results

Descriptive Statistics

Table 1 depicts the descriptive statistics of student-level variables presented in this study as a reference. Students in the study had an average 83% AET and an average DB of 9%. With regards to student-level academic achievement, the mean words read correct per minute was 23.63 words, and the math digits correct per minute average was 1.52 digits. Table 2 displays the descriptive statistics of the school-level variables as a reference. Schools had an average AET of 83% and DB of 9%. The average school FRL percentage was 61% and the average school percentage of white/non-Hispanic students was 56%. Finally, the school average for math was 12.19 digits correct per minute and for reading was 23.33 words read correct per minute. When looking at Title I status (as defined as $\geq 80\%$ FRL), 29% percent of the schools ($n= 16$) were Title I schools and 71% were non-Title I ($n= 40$). Further, 34% of schools ($n= 19$) had less than 30% of white/non-Hispanic students, 32% of schools ($n=18$) had 30-54% of white/non-Hispanic students, and 34% of schools ($n=19$) had 55% or more white/non-Hispanic students.

Research Question 1

Descriptive Statistics were used as the primary means to analyze the data in order to answer the first research question: *What is the Kindergarten, First, and Second grade normative data for Academic Engaged Time (AET) and Disruptive Behavior (DB)?*

Table 3 presents the results for the total sample ($n=6,592$). Of the 6,592 students in the study that were observed for two minutes each using momentary time sampling for on-task behaviors, the average percent of intervals on-task or academically engaged was 83%. At the grade level, Kindergarten and First grade students in the study were observed

to be on-task for 83% of the observed intervals ($SD= .20$), while Second grade students were observed to be on-task slightly less with 80% of observed intervals ($SD= .24$). The median on-task percentage across the total sample and all grade levels was 92%. Of the total sample, 2,444 students were on task 100% of time (37% of the sample; mode= 1.00). Thus, the AET data was associated with a negatively skewed distribution. Table 4 displays DB observed of the total sample and is further broken down by grade level. The average percentage of DB observed for the total sample was 9% ($SD= .05$). For Kindergarten students, the mean disruptive behaviors observed was 10% ($SD= .16$) and for First and Second grade students 9% DB ($SD= .14$). Of the total sample, 3,803 students (51% of all participants) were observed to be disruptive 0% of the time (mode=0), which demonstrates a positively skewed distribution of scores.

Table 5 represents district ($n=6$) mean AET of 82% ($SD= .03$), the mean school ($n=61$) AET of 83% ($SD= .06$), and the mean classroom ($n=318$) AET of 83% ($SD= .09$). Table 6 displays the mean district ($n=6$) DB of 10% ($SD= .02$), the mean school ($n=61$) DB of 9% ($SD= .05$), and the mean classroom ($n=318$) DB of 10% ($SD= .07$). In order to further explore the variability within the data, the ranges for percentage of AET and DB are provided in Table 5 and 6. When looking at the AET percentages across regions, the AET percentage ranged from 81-86% with the Puget Sound region averaging 81% of AET, and Mesa, Arizona region averaging 86% AET. When looking at all of the participating districts in the study ($n=6$), the range of AET is wider with a minimum average of 77% and maximum 86%. Of all the participating schools in the study ($n=61$), the AET percentages ranged from 61-91%, indicating greater variability across schools than between districts. Finally, at the classroom level ($n=318$), the percentage of AET ranged from 50-99%, indicating greater variability across

classrooms than between schools. When looking at disruptive behaviors, Mesa region schools had an average of 7% observed DB and the Puget Sound area had 10% observed DB. Across all participating districts ($n=6$), DB percentages ranged from 7-14%, and participating schools ($n=61$) had DB percentages ranging from 3-23%. The range is wider when looking at all of the participating classrooms ($n=318$), which had DB percentages that ranged from 0-37%. Table 7 presents the AET descriptive statistics for each participating school district. The mean district AET ranged from 77% to 86% percent. The average AET for each school district was as follows: Federal Way-79%, Lake Washington-82%, Mesa-86%, Mukilteo-83%, North Kitsap-77%, and Renton-84%.

Table 8 represents the percentile ranks for the total sample and grade levels in order to better understand the relative position of a score in relation to all other scores. Across all participants and grade levels, observed AET of 92% represents the 50th percentile. Academic engaged time (AET) at the 5th percentile was 42% and at the 25th percentile 75%. Given the positively skewed distribution of AET behaviors, percentile ranks of AET at the 75th, 90th, and 95th percentiles were 100% on-task behaviors. To better interpret student behavior in the classroom, off-task (OFT) percentiles are also represented in Table 8. Similar to disruptive behaviors, the data for off-task behaviors are negatively skewed. For the total sample, OFT behavior at the 50th percentile is 8%, at the 75th percentile is 25%, and at the 95th percentile is 58%. DB occurred less across all observations. Disruptive behavior at the 50th percentile for the total sample was 0%, at the 75th percentile 17%, and at the 95th percentile 42%.

Table 9 presents the results for the bivariate correlations calculated at the student level for AET, DB, and reading and math performance. Preliminary analysis showed the relationships between math and reading performance and AET/DB and academic performances

to be monotonic, as assessed by visual inspection of a scatterplot. A Spearman's rank-order correlation was run to assess the relationship between math (digits correct per minute) and reading performance (words read correct per minute). There was a strong positive correlation between reading and math performance, $r_s(6445) = .755, p < .001$. Also, there was a positive association between AET and reading performance, $r_s(6157) = .037, p = .004$. Further, results indicated that there was a statistically significant relationship between AET and math performance, $r_s(6154) = .041, p = .001$. DB was negatively correlated with AET and academic performances. As shown in Table 9, there was a moderate negative correlation between AET and DB, $r_s(6590) = -.435, p < .001$. Further, there was a statistically significant relationship between DB and reading performance, $r_s(6165) = -.043, p = .001$, and DB and math performance, $r_s(6163) = -.057, p < .001$.

Table 10 presents the bivariate correlations calculated for school-level data for the following variables: AET, DB, Reading Fluency (words correct per minute), Math Total Problem Correct, Math Digits Correct Per Minute, Off-Task (OFT), School-Wide Reading Performance (school mean of students meeting state-wide reading standards), Free and Reduced Lunch Status, Special Education Status, English Language Learner (ELL) Status, and Number of Students Per Classroom. Preliminary analysis showed the relationships between variables to be monotonic, as assessed by visual inspection of a scatterplot.

As presented in Table 10, there was a strong negative correlation between AET and DB, $r_s(59) = -.68, p < .001$. There was no significant correlation between Reading Fluency and AET, $r_s(59) = .10, p = .452$, and Reading Fluency and DB, $r_s(59) = -.20, p = .132$. In contrast, there was a moderate correlation between AET and Math Total Problems Correct, $r_s(59) = .30, p = .019$, and a negative moderate correlation between DB and Math Total Problems Correct, r_s

(59) = $-.31, p = .016$. There was a strong positive correlation between Reading Fluency and Math Total Problems Correct, $r_s(59) = .75, p < .001$. Similarly, there were significant correlations between AET and Math Digits Correct Per Minute, $r_s(59) = .27, p = .039$, and a negative correlation between DB and Math Digits Correct Per minute, $r_s(59) = -.29, p = .026$. Further, there was a strong positive correlation between Reading Fluency and Math Digits Correct Per Minute, $r_s(59) = .76, p < .001$, and Math Total Problems Correct and Math Digits Correct Per Minute, $r_s(59) = .99, p < .001$.

Results indicated that there were no significant correlations between Special Education Status and any other of the variables. ELL Status had a moderate significant negative association with School-Wide Reading, $r_s(59) = -.47, p < .001$, and a positive moderate association with FRL Status, $r_s(59) = .35, p = .006$. There was a significant association between Reading Fluency and ELL Status, $r_s(59) = -.28, p = .031$, and Math Total Problems Correct and ELL Status, $r_s(59) = -.27, p = .024$. There were no significant correlations between ELL Status and AET, DB, OFT, or Special Education Status. Class Size (number of students per classroom) was significantly positively correlated with AET, $r_s(59) = -.43, p < .001$, and negatively correlated with DB, $r_s(59) = -.41, p = .001$. Further, there were positive significant associations with Math Total Problems Correct and Class Size, $r_s(59) = .47, p = .001$, and Math Digits Correct Per Minute and Class Size, $r_s(59) = .48, p < .001$. Results indicated that there was a significant negative association between OFT and Class Size, $r_s(59) = -.43, p < .001$, and between FRL Status and Class Size, $r_s(59) = -.48, p = .268$. Finally, there were no significant associations between Reading Fluency and Class Size, Special Education Status, or ELL Status.

Research Question 2

To answer the second research question, *What is the variability in normative data across classrooms, schools and districts?*, distribution plots were created to demonstrate the variability within regions, districts, and schools. In Figure 3, a histogram of school means was plotted. The majority of schools (49 out of 59) had AET means percentages between 75-90%, two schools had AET percentages between 60-70%, and five schools had AET percentages between 90-95%. Figure 4 presents the frequency of classroom means ($n=318$) of AET. Approximately 60% (193 out 318 classrooms) of all the classrooms observed had AET between 80-94%. Twelve classrooms had AET percentages between 50-64% and eighteen classrooms had 95-99% of observed AET. To represent the frequency of mean AET within districts, Figures 5 through 10 provides histograms of the school means within each district in the study. A visual analysis of each district's histogram indicates that there is a range of variability within each district. For example, the Federal Way School District, which had 13 schools participating in the study, had more variability in AET than compared to the Mesa School District, which had 20 schools participating in the study. To further analyze the variability of AET at the region, district, and school levels, Figures 11 through 13 present mean AET and DB with corresponding 95% Confidence Intervals bars. Figure 11 depicts the means with corresponding interval bars for each region; a visual examination reveals that the confidence bars do not overlap, which indicates that there is a significant difference between regions. Further, in Figure 12, which depicts district means, an examination of the corresponding confidence interval bars suggests that there are significant differences within that unit of analysis. Figure 13, which presents all of the classroom means with corresponding confidence interval bars depicts that there is a wider range of variability and significant

differences between classrooms average AET. To examine variability, a visual analysis was conducted by comparing the spread of scores between regions, districts, and schools. Results of the spread of scores of AET and DB indicated that there is more variability at the classroom level than between schools, and more variability at the school level than between districts.

Research Question 3

In order to address the final research question, *To what extent are school-level variables (percent of students qualifying for Free and Reduced Lunch [FRL], percent of white/non-Hispanic students, and academic achievement data) associated with school-level norms regarding Academic Engaged Time (AET) and Disruptive Behavior (DB)?*, linear and logistic regressions were conducted for AET and DB.

Table 11 presents the linear and logistic regression analyses of AET that were adjusted for regional affects and accounted for clustering within districts. Results of the linear regression showed that there were no statistically significant associations with Title I status, % of white/non-Hispanic students, or math and reading achievement to AET. Note that there was marginally significant result in which schools with 30-55% white/non-Hispanic populations were observed to have a 3.7% increase in AET relative to schools with <30% ($p = .070$). However, when analyzing AET in the logistic regression, Title I status and % of white/non-Hispanic students were each significantly associated with AET ($p < .05$). Title I status was associated with lower AET and compared to the reference group of non-Title I schools, Title I schools was associated with a significantly 0.26 lower odds of having high AET ($OR(95\%CI) = .26, (.11, .65), p = .004$). Moreover, school white/non-Hispanic percentage was found to be significantly associated with higher odds of high AET (30%-55% versus <30% white/non-Hispanic %

(*OR (95%CI)*= 2.78 (1.01, 7.62), *p* = .048). Given the distribution of AET with a large proportion of schools having high average AET, it is believed that the logistic regression is the better model to meaningfully analyze the school-level variables and their associations between higher engagement and lower engagement schools.

Table 12 presents the linear and logistic regression analyses of DB that were adjusted for regional affects and accounted for clustering within districts. The linear regression for each school-level predictor was not significantly associated with DB. When the DB outcome was analyzed using a logistic regression, academic achievement was associated with DB (*p* < .05). For math, an increase in one digit correct per minute was associated with a 0.49 times lower odds of high DB (*OR(95%CI)*= .49 (.26, .93), *p* = .028). Similarly, for reading, an increase in 10 words read correct per minute was significantly associated with a .084 times lower odds of high DB (*OR(95%CI)*= .84 (.80, .88), *p* = .023).

Chapter 5: Discussion

The main purpose of this study was to establish normative benchmarks for student academic engagement in early elementary classrooms, which was measured by directly observing students' on-task, off-task, and disruptive behaviors in the classroom.

Establishing normative benchmarks of academic engagement in early elementary classrooms could provide educators and researchers valuable information by helping them better understand norms of academic engaged time. By establishing normative benchmarks at the regional, district, and school levels, educators and researchers can better understand early elementary students' classroom behavior and use these observational data as a means to establish classroom and individual student goals and monitor whole class and individual student progress over time, which are vital ingredients for learning. Given the federal mandates that require schools to adequately monitor students' behavioral functioning, more research is needed to better understand behavioral trends in K-12 education in order to effectively monitor and interpret student, classroom, school, and district behavioral functioning. Another purpose of this study was to examine associations of school-level variables to behavioral engagement. Such research can better inform educators as to what school characteristics are associated with higher or lower behavioral engagement. This provides researchers and educators an additional lens in interpreting behavioral engagement in elementary schools and could allow for an increase of supports for schools that may be more at-risk for behavioral disengagement.

The results from this study indicate that the average academic engaged time in early elementary classrooms as measured through systematic direct observational

methods is approximately 80% with students displaying off-task behavior about 20% of the time, meaning for every hour of instruction students as a whole spend roughly 48 minutes academically engaged and 12 minutes disengaged. Disruptive behaviors that impact classroom learning were observed to occur approximately 10% of the time in early elementary classrooms. On-task, off-task, and disruptive behavior means were consistent across Kindergarten, First, and Second grades. Results also suggested that the lower the unit or level of analysis, the more variability was found. For example, while overall district academic engagement was similar, when looking more closely at school and classroom engagement, there was a wider range of scores, which suggests that this observational data could serve as an indicator of overall academic engaged time that could be compared to district and grade level norms meaningfully. Normative benchmark data on early elementary students' academic engaged time was relatively unknown prior to this study because few studies of this magnitude have utilized systematic direct observation of students' classroom behavior.

Results from this study further indicated that there are significant associations between school-level variables and observed academic engagement and disruptive behavior. Bivariate correlations indicated that free and reduced lunch status, English language learner status, and class size all had significant associations with behavioral engagement. In addition, the math curriculum-based measurement used in this study was significantly correlated to behavioral engagement, whereas the reading fluency curriculum-based measurement was not significantly associated with behavioral engagement. This is surprising since prior research has shown that reading fluency measures are sensitive predictors of academic achievement. In the limitations section in

the chapter, further examination of the reading fluency measure used in this study is discussed. One explanation for this finding is that responsiveness to early reading instruction is more dependent on school readiness with regard to language and early literacy skills than academic engagement in instruction, whereas math performance is more dependent on early learning experiences in the classroom in the early elementary years. These results, however, need to be interpreted with discretion, as they were not conducted using multi-level models to account for nesting of students within classrooms and classrooms within schools. That being said, the correlations provide descriptive data that have implications for future multi-level research on behavioral engagement.

Consistent with the hypotheses underlying this study, results of the logistic regression indicated that there are school characteristics that are significantly associated with school-level AET and DB. Socio-economic factors, as measured by Free and Reduced Lunch percentages (Title I status), and school race demographics were associated with AET. Specifically, Title I schools were associated with lower AET, and higher white/non-Hispanic school populations were associated with higher AET. It should be noted that this is not a causal study, and instead suggests relationships between the variables. Further, prior educational research has long demonstrated that Title I schools, which tend to be more racially and ethnically diverse, are often underserved and many complex factors have long contributed to achievement gaps between high and low socio-economic groups (Shin, 2005). The results of this study further confirm a wide body of research that socio-economic factors still persist as strong indicators in a multitude of school performance outcomes. While socio-economic and race factors were significantly associated with academic engagement, they were not associated with an

increase in disruptive behaviors. It is difficult to discern whether the reduced association of DB with the socio-economic variable is due to student-level factors or context/environmental factors associated with classroom or school settings (e.g., classroom management, school funding, school cultural expectations of students). Future research will want to explore the individual- and contextual-level factors that impact student engagement in the classroom, particularly in schools serving a large proportion of students from economically disadvantaged settings. School academic achievement, as measured by curriculum-based measurements in math and reading, were associated with the disruptive behavior outcome. Finally, schools in the high disruptive behavior group were associated with lower math and reading achievement. These results are not surprising given that problematic behaviors are incompatible with engagement and learning in the classroom.

Limitations

One of the limitations of this study is that academic engaged time is solely measured by observable, overt behaviors and does not account for unobservable engagement such as intrinsic feelings, moods, thoughts, or beliefs. In addition, interval time sampling of student behaviors does not allow us to collect data on the quality of behavior, such as the frequency, intensity, or duration of on-task, off-task, and disruptive behaviors. Instead, the descriptive statistics outlined in this study are meant to provide a broad view of student classroom behavior. Further, one of the limitation of AET as a metric is that it does not account for teacher behavior, and their effect on engagement. Teacher behavior affects the degree to which students are active participants in the learning experience (Berliner, 1984; Skinner, 2004; Skinner, Pappas, & Davis, 2005).

Skinner et al. (2004) noted that students' engagement can be increased with active accurate academic (AAA) engagement, which involves ample opportunities to respond. For example, teachers can increase students' opportunities to respond in the classroom (e.g., choral responses, technological anonymous responses in surveys), which could then affect engagement. Since behavior is related to the environment, more research is needed to see what other variables contribute to student engagement.

Another limitation of this study was the potential for reactivity effects caused by the presence of the observers. However, observers attempted to acclimate students to their presence by spending time in the classroom and administering curriculum-based measurements with each student prior to conducting the direct observation. Further, it should be noted that the author recognizes that classrooms are complex environments where student behavior is based on a multitude of factors (Ratcliff et al., 2010). Measuring only the discreet categories of on-task, off-task, and disruptive behaviors may over-simplify and miss the complexity of student-teacher interactions and how that relates to overall classroom norms. In addition, developing normative benchmarks requires a large population sample, geographical spread, and development of reliability and validity of the measures used. While the sample size for this study provides a large amount of data and relatively representative sample ethnically, observational data was collected only in the Puget Sound and Mesa, Arizona regions and more research is needed to better understand academic engaged time in other regions of the United States. Another limitation of this study is the cut-off scores used to determine the binary categories for the logistic regressions. Since there was no prior research on quantitative categories of engagement, the author constructed binary categories by examining the

descriptive data. Results from the logistic regression are not intended to provide causal inference for behavioral engagement, but instead are useful in initially examining the relationships in data that has not been tested before; results should be interpreted with caution as the design of this study cannot draw any causal associations. In addition, the reading fluency measure used in this study is considered a valid and reliable measure of reading performance in First and Second grades, but to a far lesser extent in Kindergarten. A more appropriate measure for measuring reading performance in Kindergarten would be a letter-sound fluency or a letter-naming fluency curriculum-based measurement, which are appropriate emergent reading skills for this age group. The reading fluency measure was included in the larger study to examine growth rates over time in a longitudinal study, but it is believed that using the reading fluency measure for this descriptive study was not sensitive enough in capturing reading performance. Finally, another limitation is that this study did not utilize multi-level models to investigate variables at different levels that impact student academic engagement (i.e., district and class levels of analysis). While the present study's purposes were to establish normative benchmarks and examine school-level variables associated with behavioral engagement, future research should utilize multi-level models to better understand the factors that explain academic engagement in the classroom.

Future Implications

Future research should explore the extent to which observational data of students' behavioral engagement are associated with other teacher ratings of students' behavioral, cognitive, and affective engagement. Additional criterion related validity work should also be done to link on-task, off-task, and disruptive behaviors with other relevant

educational variables. Further, this study indicates that students are not engaged 1 out of every 5 minutes in the classroom, which is roughly 12 minutes out of every instructional hour. Increasing academic engaged time could be accomplished by identifying classrooms and schools in need and SDO data could help educators attain a deeper understanding of where precious instructional time is lost. If a classroom is identified through SDO data as being in need of supports, then that teacher could be provided effective proactive classroom management strategies to increase academic engaged time. Similarly, a school could be identified through SDO data as needing a more effective building-wide approach to increase academic engaged time. Effective building-wide approaches, such as implementing a social-emotional curriculum or a PBIS model, could be implemented at the Tier 1 general education level when a school is identified as needing to increase or maintain their academic engaged time.

It can be argued that one of teachers' main responsibilities is to engage students in academic learning activities. The quality of the curriculum will dictate the degree to which students learn when they are engaged in instruction, as well as the instructional strategies teachers utilize in order for instruction to be effective. In the current educational environment, teacher evaluations are increasingly tied to "value-added" measures. It will be important to develop tools that are sensitive to local norms in order to ensure that teachers are evaluated fairly and accurately. Currently, teacher evaluations of classroom management and student engagement are more often tied to rubrics (e.g., Danielson Framework) or observations that are not systematic in nature. While these evaluation methods are often considered useful and valid, it is argued that developing reliable and valid benchmarks of academic engaged time that are sensitive to school-level

factors would further increase the validity of teacher evaluation systems. Further, academic engaged time has strong evidence for being a necessary condition for learning in the classroom and teachers and administrations find academic engaged time in classrooms a highly acceptable, social valid outcome (Appleton et al. 2006; Christenson et al., 2008; Finn & Zimmer, 2012; Fredricks et al., 2004). Normative benchmarks of academic engaged time can assist in data based decision making in a RtI/MTSS system when implementing school and classroom wide supports around classroom management and positive behavioral supports. Effective instructional strategies could be implemented and monitored through systematic observational data in order to increase academic engaged time and reduce disruptive behavior.

The feasibility of conducting district-wide and school-wide systematic direct observations should be explored. While systematic observations have been shown to be reliable and valid measures of behavioral engagement, much of the current observational data conducted in schools is for individuals suspected of having a disability (Brown-Chidsey & Andren, 2013). Implementing observational systems for universal screening and progress monitoring at the Tier 1 and Tier 2 service delivery levels may not be a feasible endeavor as it requires staff time and training to effectively implement this data collection procedure. However, current behavioral student measurements for classrooms and schools tend to be reactive in nature (e.g., detention and suspension rates, discipline referral data), and it is argued that more proactive universal measures of behavioral engagement are needed in order to effectively interpret and measure classroom and school-wide behavioral engagement. More research is needed to better understand the

social validity of using observational measures of student engagement in classrooms and schools.

Finally, the GEE statistical approach utilized in this study is not a common method used in psychological and educational research. A traditional generalized linear model could not be used because the data violated the assumption of normality and independence. Because the AET and DB outcome data used in this study was not normally distributed, and there was dependence between variables as found in the bivariate analysis, the GEE approach was utilized in order to effectively address these limitations. The GEE approach is becoming increasingly popular to address non-normally distributed data, but this method is most often used in the biostatistics field. Future research exploring the GEE method in educational and psychological research should be examined as this approach addressed some of the common barriers encountered in educational research, which often deals with nested and multicollinearity.

Conclusion

Clearly more research is needed to develop robust normative benchmarks of students' academic engaged time. To this author's knowledge, the present study is the largest to date to conduct direct observation of students' behavioral academic engagement in the classroom, namely on-task, off-task, and disruptive behaviors. The findings from this study could provide useful information to develop classroom or school-wide goals, as well as provide a reference point to interpret student engagement and evaluate response to intervention(s) that are targeting increased academic engagement. These data could also be used to inform administration and teacher

evaluation efforts, as academic engagement is one of the primary goals of creating an effective learning environment for all students.

References

- Adelman, H. & Taylor, L. (2012). Mental health in schools: Moving in new directions. *Contemporary School Psychology, 16*, 9-18.
- Agnoff, W.H. (1989). Scales, Norms, and Equivalent Scores. Princeton, NJ: Educational Testing Service.
- Alessi, G. (1980). Behavior observation for the school psychologist: Responsive-discrepancy model. *School Psychology Review, 9*, 31-45.
- Appleton, J.J., Christenson, S.L., & Furlong, M.J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools, 45*, 5, 369-386.
- Appleton, J. J., Christenson, S. L., Kim, D., Reschly, A. L., & Motivation. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*, 5, 427-445.
- Bailey J.S, Burch M.R. (2002). Research methods in applied behavior analysis. Thousand Oaks, CA: Sage.
- Baugous, K., & Bendery, S. (2000). Decreasing the amount of classroom disruptions in order to increase the amount of time on task in elementary students. Dissertation retrieved from ERIC. <http://eric.ed.gov/?id=ED443554>
- Becker, K.E. & Domitrovich, C.E. (2011). The conceptualization, integration, and support of evidence-based interventions in the schools. *School Psychology Review, 40*, 4, 582-589.

- Benedict, A. E., Thomas, R. A., Kimerling, J., & Leko, C. (May 01, 2013). Trends in Teacher Evaluation: What Every Special Education Teacher Should Know. *Teaching Exceptional Children, 45, 5, 60-68.*
- Betts, J. E., Appleton, J. J., Reschly, A. L., Christenson, S. L., & Huebner, E. S. (2010). A study of the factorial invariance of the Student Engagement Instrument (SEI): Results from middle and high school students. *School Psychology Quarterly, 25, 2, 84-93.*
- Berliner, D.C. (1984). The half-full glass: A review of research on teaching. In P.L. Hosford (Ed.), *Using what we know about teaching (51-85)*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Briesch, A.M., Chafouleas, S.M., & Riley-Tillman, T.C. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison study of systematic direct observation and direct behavior rating scale. *School Psychology Review, 39, 3, 408-421.*
- Brown-Chidsey, R. & Steege, M. (2005). *Response to intervention: Principles and strategies for effective practice*. New York: Guilford Press.
- Brown-Chidsey, R. & Andren, K. (2013). *Assessment for intervention: A problem-solving approach*. New York: Guildord Press.
- Burns, M. K., Jacob, S., & Wagner, A. R. (2008). Ethical and legal issues associated with using response-to-intervention to assess learning disabilities. *Journal of School Psychology, 46, 263-279.*
- Carter, C. P., Reschly, A. L., Lovelace, M. D., Appleton, J. J., & Thompson, D. (2012). Measuring student engagement among elementary students: Pilot of the Student

Engagement Instrument--Elementary Version. *School Psychology Quarterly: The Official Journal of the Division of School Psychology*, American Psychological Association, 27, 2, 61-73.

Christenson, S.L., Whitehouse, E.M., & VanGetson, G.R. (2008). Partnering with families to enhance students' mental health. In B. Doll & J. Cummings (Eds) *Transforming School Mental Health Services: Population Based Approaches to Promoting the Competency and Wellness of Children*, (pp.69-102).

Cohen, L., Manion, L., & Morrison, K. (2003). *Research methods in education*. London: Routledge Falmer.

Connell, J. P., Spencer, M. B., & Aber, J. L. (1994). Educational risk and resilience in African-American youth: Context, self, action, and outcomes in school. *Child Development*, 65, 493-506.

Conrad, H.S. Norms. In C.W. Harris (Ed.), *Encyclopedia of Educational Research, Revised Edition*. New York: MacMillian, 1950.

Cook, B. G., & Cook, S. C. (2011). Unraveling evidence-based practices in special education. *Journal of Special Education*, 47(2), 71-82. doi: 10.1177/0022466911420877

Cook, C.R., Browning-Wright, D., Gresham, F.M., & Burns, M.K. (2010). *Transforming School Psychology in the RTI Era: A Guide for Administrators and School Psychologists*. Horsham, PA: LRP Publications.

Cook, B. G., & Odom, S. L. (2013). Evidence-Based Practices and Implementation Science in Special Education. *Exceptional Children*, 79(2), 135-144.x

- Cook, B.G. & Schirmer, B.R. (2006). An overview and analysis of the role of evidence-based practices in special education. In B. Cook & B. Schirmer (Eds), *What is Special About Special Education?* (pp.165-174). Austin, TX: Pro-Ed.
- DeSocio, J., & Hootman, J. (2004). Children's mental health and school success. *Journal Of School Nursing*, 20(4), 189-196. doi:10.1177/10598405040200040201
- Deno, S. (2005). Problem-solving assessment. In R. Brown-Chidsey (Ed.), *Assessment for intervention: A problem-solving approach*. New York: Guilford Press (pp. 10-40).
- Doll, B., & Cummings, J. A. (2008). Transforming school mental health services: Population-based approaches to promoting the competency and wellness of children. Thousand Oaks, CA: Corwin Press.
- Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions, *Child Development*, 82, 1, 405-432.
- Finn, J.D., & Zimmer, K.S. (2012). Student Engagement: What is it? Why does it matter? In S.L. Christenson, A.L. Reschly, & C. Wylie (Eds.), *The handbook of research on student engagement*. New York, NY: Springer (pp.97-130).
- Fredricks, J. A. (2011). Engagement in School and Out-of-School Contexts: A Multidimensional View of Engagement. *Theory into Practice*, 50, 4, 327-335.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74, 1, 59-109.

- Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., Mooney, K., & Regional Educational Laboratory Southeast (ED). (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. Regional Educational Laboratory Southeast at SERVE Center. University of North Carolina at Greensboro, 915 Northridge Street, Greensboro, NC 27403. Tel: 800-755-3277; e-mail: RELSoutheast@serve.org; Web site: <http://www.serve.org/RELSE.aspx>.
- Furrer, C., & Skinner, E. (2003). Sense of Relatedness as a Factor in Children's Academic Engagement and Performance. *Journal of Educational Psychology*, 95, 1, 148-62.
- Gresham, F.M. (1981). Social skills training with handicapped children: A review. *Review of Educational Research*, 51, 139-176.
- Gresham, F., Reschly, D., & Shinn, M. (2010). RTI as a driving force in educational improvement: Research, legal, and practice perspectives. In H. Walker & M. Shinn (Eds) *Interventions for Achievement and Behavior Problems in a Three-Tier Model Including RTI*, (pp. 47-78).
- Haynes, S.N. & Wilson, C.C. (1979). Behavioral assessment: Recent advances in methods, concepts, and applications. San Francisco: Jossey-Bass.
- Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. *Best Practices in School Psychology*, 4, 993-1006.
- Hoagwood, K. (2003-04). Evidence-based practice in child and adolescent mental health: Its meaning, application and limitations. *Emotional & Behavioral Disorders in Youth*, 4, 7-8.

- Hosp, M., Hosp, J., & Howell, K. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.
- Individuals with Disabilities Education Improvement Act of 2004, PL 108-466, 20 U. S. C. §1400, H. R. 1350.
- Finn, J. D. (1993). *School engagement and students at risk*. Washington, DC: National Center for Education Statistics.
- Finn, J.D. (1989). Withdrawing from school. *Review of Educational Research*, 59, 117-142.
- Finn, J.D. & Rock, D.A. (1997). Academic success among students at risk for school failure. *The Journal of Applied Psychology*, 82, 2, 221-234.
- Fisher, C.W. & Berliner, D.C. (1985). *Perspectives on instructional time*. New York: Longman.
- Frey, A., Lingo, A., and Nelson, C.M. (2010). Implementing Positive Behavior Support in Elementary Schools. In H. Walker & M. Shinn (Eds) *Interventions for Achievement and Behavior Problems in a Three-Tier Model Including RTI*, (pp. 397-434).
- Gettinger, M. & Seibert, J.K. (2002). Best practices in increasing academic learning time. In A. Thomas & J. Grimes (Eds.), *Best practices in School Psychology*, IV, 773-787. Bethesda, MD: National Association of School Psychologists.
- Johns, E. B. H., Crowley, P., & Guetzloe, E. (2008). Engaged Time in the Classroom. *Focus on Exceptional Children*, 41, 4, 1.

- Kaplan, D.S., Peck, B.M., & Kaplan, H.B. (1997). Decomposing the academic failure-dropout relationship: A longitudinal analysis. *Journal of Educational Research*, 90, 6, 331-343.
- Kauchak, D. P., & Eggen, P. D. (1993). *Learning and teaching: Research-based methods*. Boston: Allyn and Bacon.
- Klein, R.D. (1979). Modifying academic performance in the grade school classroom. *Progress in Behavior Modification*, 8, 293-321.
- Klem, A.M. & Connell, J.P. (2004). Relationships matter: Linking teacher support of student engagement and achievement. *Journal of School Health*, 74, 7, 262-273.
- Kratochwill, T.R., Hoagwood, K.E., Kazak, A.E., Weisz, J.R., Hood, K., Vargas, L.A. and Banez, G.A. (2012). Practice-based evidence for children and adolescents: Advancing the research agenda in schools. *School Psychology Review*, 41, 2, 215-235.
- Hardin, J.W. and Hilbe, J.M. (2002). Generalized Estimating Equations. Retrieved from <http://www.ehlib.com>
- Hintze, J.M. (2005). Psychometrics of direct observation. *School Psychology Review*, 34, 507-519.
- Ikeda, M., Neesen, E., & Witt, J. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology*, 5, 103-114. Bethesda, MD: National Association of School Psychologists.
- Kuh, G.D. (2003). What we're learning about student engagement from NSSE: Benchmarks for effective educational practices. *Change*, 35, 2, 24-32.

- Lane, K., Oakes, W., & Menzies, H. (2010). Systematic Screenings to Prevent the Development of Learning and Behavior Problems: Considerations for Practitioners, Researchers, and Policy Makers. *Journal of Disability Policy Studies*, 21, 3, 160-172.
- Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice*. San Francisco, CA: Jossey-Bass.
- Maag, J.W. & Katsivannis, A. (2010). Early intervention programs for children with behavior problems and at risk for developing antisocial behaviors: Evidence- and research-based practices. *Remedial and Special Education*, 31, 6, 464-475.
- Marks, H. M. (2000). Student Engagement in Instructional Activity: Patterns in the Elementary, Middle, and High School Years. *American Educational Research Journal*, 37, 1, 153-84.
- Merrell, K. W. (2010). *Behavioral, social, and emotional assessment of children and adolescents*. New York: Routledge.
- Mosher, R. & MacGowan, B. (1985). *Assessing student engagement in secondary schools: Alternative conceptions, strategies of assessing, and instruments*. Madison, WI: University of Wisconsin Research and Development Center. Retrieved from ERIC database (ERIC document reproduction service no. ED 272812).
- National Association of School Psychologists (2006). Retrieved from <http://www.nasponline.org/resources/handouts/sbmhservices.pdf>
- National Center for Education Evaluation and Regional Assistance/US Department of Education Institute of Education Sciences (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*.

Coalition for Evidence-based Policy. Retrieved from:

<http://www2.ed.gov/rschstat/research/pubs/rigorousetid/rigorousetid.pdf>

Newmann, F.M. (1992). Student engagement and achievement in American secondary schools. New York: Teachers College Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115, § 1425, 20 U.S.C. §§ 6301 et seq. (2002).

Nystrand, M. & Gamoran, A. (1991). Instructional discourse, student engagement, and literature achievement. *Research in the Teaching of English*, 25, 261-290.

Pintrich, P., Wolters, C., & Baxter, G. (2000). Assessing metacognitions and self-regulated learning. In G. Schraw & J. Impara (Eds.), *Issues in Measures of Metacognition* (pp. 43-97). Lincoln NE: Buros Institute of Mental Measurement.

Power, T., Mautone, J., & Ginsburg-Block, M. (2010). Training School Psychologists for prevention and interventions in a three-tier model. In H. Walker & M. Shinn (Eds) *Interventions for Achievement and Behavior Problems in a Three-Tier Model Including RTI*, (pp. 151-174).

Prater, M. (1992). Increasing Time-on-Task in the Classroom. *Intervention in School and Clinic*, 28, 1, 22-27.

President's New Freedom Commission on Mental Health (2003). *Achieving the promise: Transforming mental health care in America: Executive summary: Final Report*.

Retrieved from <http://permanent.access.gpo.gov/lps38967/ExecSummary.pdf>

Ratcliff, N. J., Jones, C. R., Costner, R. H., Savage-Davis, E., & Hunt, G. H. (2011). The Elephant in the Classroom: The Impact of Misbehavior on Classroom Climate. *Education*, 131, 2, 306-314.

- Rathvon, N. (2008). *Effective school interventions: Evidence-based strategies for improving student outcomes*. New York: Guilford Press.
- Reith, H., & Evertson, C. (1988). Variables related to effective instruction of difficult-to-teach children. *Focus on Exceptional Children*, 20, 1–8.
- Reschly, A.L. & Christenson, S.L. (2012). Moving from “context matters” to engaged partnerships with families. *Journal of Educational & Psychological Consultation*, 22, 62-78.
- Sackett, D.L., Rosenburg, W.M., Gray, J.A., Haynes, R.B., & Richardson, W.S. (1996). Evidence-based medicine: What it is and isn't. *British Medical Journal*, 312, 71-72.
- Salvia, J. & Ysseldyke, J.E. (2004). *Assessment in special and inclusive education*. Boston: Houghton Mifflin.
- Seeley, J.R., Rohde, P., Backen Jones, L. (2010). School based prevention and intervention for depression and suicidal behavior. In Shinn, M. R., & Walker, H. M. (Eds), *Interventions for achievement and behavior problems in a three-tier model including RTI*. Bethesda, MD: National Association of School Psychologists (pp. 363-396).
- Sattler, J. M., & Hoge, R. D. (2006). *Assessment of children: Behavioral, social, and clinical foundations*. San Diego: J.M. Sattler.
- Schrader, W.B. (1960). Norms. In C.W. Harris (Ed.), *Encyclopedia of Educational Research*, 3rd Edition. New York: MacMillian, 1960.
- Shapiro, E.S. (2011). *Academic skills problems: Direct assessment and intervention*. New York: Guilford Press.

- Shapiro, E.S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York: Guilford Press.
- Shapiro, E.S., Hilt-Panahon, A., & Gischlar, K.L. (2010). Implementing proven research in school-based practices: Progress monitoring within a response-to-intervention model. In H. Walker & M. Shinn (Eds) *Interventions for Achievement and Behavior Problems in a Three-Tier Model Including RTI*, (pp. 175-192).
- Shapiro, E. S., & Kratochwill, T. R. (2000). *Behavioral assessment in schools: Theory, research, and clinical foundations*. New York: Guilford Press.
- Shapiro, E.S. & Heick, P. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavior/emotional problems. *Psychology in the Schools*, 41, 551-561.
- Shulman, L.S. (2002). Making differences: A table of learning. *Change*, 34, 36-44.
- Sinclair, M.F., Christenson, S.L., Evelo, D.L., & Hurley, C.M. (1997). Dropout preventions for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children*, 65, 1, 7-21.
- Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75, 3, 417-453.
- Skinner, C.H. (2004). Academic Engagement. In A. Canter (Ed.), *Helping Children at Home and School-II* handouts from your school psychologist. Bethesda, MD: National Association of School Psychologists.
- Skinner, C.H., Pappas, D.N., & Davis, K.A. (2005). Enhancing academic engagement: Providing opportunities for responding and influencing students to choose to respond. *Psychology in Schools*, 42, 4, 389-403.

- Skinner, E., Kindermann, T., & Furrer, C. (2009). A Motivational Perspective on Engagement and Disaffection. *Educational and Psychological Measurement*, 69, 3, 493-525.
- Soukup, J. H., Wehmeyer, M. L., Bashinski, S. M., & Bovaird, J. A. (2007). Classroom variables and access to the general curriculum for students with disabilities. *Exceptional Children*, 74, 101–120.
- Spanjers, D., Burns, M., & Wagner, A. (2008). Systematic Direct Observation of Time on Task as a Measure of Student Engagement. *Assessment for Effective Intervention*, 33, 2, 120-126.
- Sprague, J., Cook, C., Browning-Wright, D., & Sadler, C. (2008). Response to intervention for behavior: Integrating academic and behavior supports. Palm Beach, FL: LRP.
- Sprick, R. & Borgmeier, C. (2010). Behavior Prevention and Management in Three Tiers in Secondary Schools. In H. Walker & M. Shinn (Eds) *Interventions for Achievement and Behavior Problems in a Three-Tier Model Including RTI*, (pp. 435-468).
- Sprick, R., Knight, J., Reinke, W., & McKale, T. (2006). Coaching classroom management: Strategies and tools for administrators and coaches. Eugene, OR: Pacific Northwest Publishing.
- Sugai, G., & Horner, R. (2002). The Evolution of Discipline Practices: School-Wide Positive Behavior Supports. *Child & Family Behavior Therapy*, 24, 23-50.
- Sugai, G., & Horner, R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, 35, 245-259.

- Tankersley, M., Harjusola-Webb, S., and Landrum, T. (2008). Using single-subject research to establish the evidence base of special education. *Intervention in School and Clinic, 44* (2), 83-90. doi: 10.1177/1053451208321600
- United States Department of Education: Office of Special Education and Rehabilitative Services Office of Special Education Program (1991). Chapter 1 policy manual: State-operated or supported programs for handicapped children. Washington D.C.
- US Public Health Service. *Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda*. Washington, DC: Department of Health and Human Services; 2000.
- Volpe, R.J. & McConaughy, S.H. (2005). Systematic direct observational assessment of student behavior: its use and interpretation in multiple settings: An introduction to the miniseries. *School Psychology Review, 34*, 4, 451-453.
- Walker, H., Severson, H., Nicholson, F., Kehle, T., Jenson, W., & Clark, E. (1994). Replication of the Systematic Screening for Behavior Disorders (SSBD) Procedure for the Identification of At-Risk Children. *Journal of Emotional and Behavioral Disorders, 2*, 2, 66-77.
- Walker, H.M., Severson, H.H., & Seeley, J.R. (2010). Universal, school-based screening for the early detection of behavioral problems contributing to later destructive outcomes. In H. Walker & M. Shinn (Eds) *Interventions for Achievement and Behavior Problems in a Three-Tier Model Including RTI*, (pp. 677-702).
- Walker, H.M. & Shinn, M.R. (2010). Systematic, evidence-based approaches for promoting positive student outcomes within a multitier framework: Moving from efficacy to effectiveness. In H. Walker & M. Shinn (Eds) *Interventions for*

Achievement and Behavior Problems in a Three-Tier Model Including RTI, (pp. 1-26).

Whitehurst, G.J. (2001). *Evidence-based education (EBE)*. Retrieved from http://ies.ed.gov/director/pdf/2002_10.pdf

Willingham, D.T. (2009). *Why don't students like school?: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. San Francisco, CA: Jossey-Bass.

Wilson, M.S. & Reschly, D. (1996). Assessment in school psychology training and practice. *School Psychology Review*, 25, 1, 9-23.

Winn, D. D., Menlove, R., & Zsiray, J. S. W. (1997). An Invitation to Innovation: Rethinking the High School Day. *NASP Bulletin*, 81, 588, 10.

Winne, P. & Perry, N. (2000). Measuring self-regulated learning. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531-566). San Diego, CA: Academic Press.

Zeger, S. & Liang, K. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 1, 121-130.

Ziegler, A. (2011). *Generalized Estimating Equations*. Springer-Verlag: New York. doi: 10.1007/978-1-4614-0499-6

Appendix

Table 1

Descriptive Statistics of Student-Level Variables

	School-Level Variables			
	<i>M</i>	<i>Mdn</i>	<i>95% CI</i>	<i>SD</i>
Academic Engaged Time (AET)	.83	.92	.82-.83	.20
Disruptive Behavior (DB)	.09	.00	.08-.10	.15
Reading (words correct per minute)	23.63	8	22.82-24.44	.24
Math (digits correct per minute)	1.52	1.13	1.49-1.56	1.45

Table 2

Descriptive Statistics of School-Level Variables

	School-Level Variables			
	<i>M</i>	<i>Mdn</i>	<i>95% CI</i>	<i>SD</i>
Academic Engaged Time (AET)	.83	.83	.81-.84	.06
Disruptive Behavior (DB)	.09	.08	.08-.11	.05
Free and Reduced Lunch Status	.61	.60	.54-.67	.24
% of White, Non-Hispanic	.56	.60	.50-.62	.24
Special Education %	.09	.09	.07-.10	.05
ELL Status %	.21	.20	.18-.25	.14
Class Size	23.39	22.60	22.55-24.23	3.25
Math (digits correct per minute)	12.19	11.64	11.09-13.30	4.13
Math (total problems correct)	8.27	7.89	7.57-9.97	2.7
% of Students Meeting Reading Standard	.74	.75	.72-.77	.10
School Reading (words read correct per minute)	23.33	21.66	20.75-25.91	9.62

Table 3

Means, Medians, Confidence Intervals, and Standard Deviations of Academic Engaged Time (AET) for Students

Total Sample and Grades	Percentage of AET				
	<i>n</i>	<i>M</i>	<i>Mdn</i>	<i>95% CI</i>	<i>SD</i>
Total Sample	6,592	.83	.92	.82-.83	.20
Kindergarten	2,770	.83	.92	.82-.84	.20
First Grade	3,443	.82	.92	.82-.83	.20
Second Grade	379	.80	.92	.78-.83	.24

Table 4

Means, Medians, Confidence Intervals, and Standard Deviations of Disruptive Behaviors (DB) for Students

Total Sample and Grades	Percentage of DB				
	<i>n</i>	<i>M</i>	<i>Mdn</i>	<i>95%CI</i>	<i>SD</i>
Total Sample	6,601	.09	.00	.09-.10	.15
Kindergarten	2,770	.10	.00	.09-.10	.16
First Grade	3,443	.09	.00	.08-.09	.14
Second Grade	380	.09	.00	.07-.10	.14

Table 5

Means, Medians, Ranges, and Standard Deviations of Academic Engaged Time (AET) for Districts, Classrooms, and Schools

	Percentage of AET			
	<i>M</i>	<i>Mdn</i>	<i>Range</i>	<i>SD</i>
Districts (<i>n</i> =6)	.82	.83	.77-.86	.03
Schools (<i>n</i> =61)	.83	.83	.61-91	.06
Classrooms (<i>n</i> =318)	.83	.84	.50-.99	.09

Table 6

Means, Medians, Ranges, and Standard Deviations of Disruptive Behaviors (DB) for Districts, Classrooms, and Schools

	Percentage of DB			
	<i>M</i>	<i>Mdn</i>	<i>Range</i>	<i>SD</i>
Districts (<i>n</i> =6)	.10	.11	.07-.14	.02
Schools (<i>n</i> =61)	.09	.08	.03-.23	.05
Classrooms (<i>n</i> =318)	.10	.08	0-.37	.07

Table 7

Means, Medians, Confidence Intervals, and Standard Deviations of Academic Engaged Time (AET) for Districts

District	Percentage of AET			
	<i>M</i>	<i>Mdn</i>	<i>95%CI</i>	<i>SD</i>
Federal Way	.79	.83	.78-.80	.22
Lake Washington	.82	.92	.81-.84	.19
Mesa	.86	.92	.86-.87	.19
Mukilteo	.83	.83	.81-.86	.18
North Kitsap	.77	.83	.75-.78	.24
Renton	.84	.92	.82-.85	.20

Table 8

Percentile Ranks of Academic Engaged Time, Off-Task Behavior, and Disruptive Behavior of Total Sample and Grade Levels

	Percentiles				
	5	25	50	75	95
<i>Total Sample (n=6,592)</i>					
Percentage of Academic Engaged Time (AET)	.42	.75	.92	1.00	1.00
Percentage of Off-Task Behavior (OFT)	.00	.00	.08	.25	.58
Percentage of Disruptive Behavior (DB)	.00	.00	.00	.17	.42
<i>Kindergarten (n=2,770)</i>					
Percentage of Academic Engaged Time (AET)	.42	.75	.92	1.00	1.00
Percentage of Off-Task Behavior (OFT)	.00	.00	.08	.25	.58
Percentage of Disruptive Behavior (DB)	.00	.00	.00	.17	.42
<i>First Grade (n=3,443)</i>					
Percentage of Academic Engaged Time (AET)	.42	.75	.92	1.00	1.00
Percentage of Off-Task Behavior (OFT)	.00	.00	.08	.25	.58
Percentage of Disruptive Behavior (DB)	.00	.00	.00	.08	.42
<i>Second Grade (n=380)</i>					
Percentage of Academic Engaged Time (AET)	.25	.67	.92	1.00	1.00
Percentage of Off-Task Behavior (OFT)	.00	.00	.08	.33	.75
Percentage of Disruptive Behavior (DB)	.00	.00	.00	.17	.33

Table 9

Spearman's Rho Correlation Matrix for Student-Level Variables

	1	2	3	4
1. Reading (words read correct per minute)	1			
2. Math (digits correct per minute)	.76**	1		
3. Academic Engaged Time (AET)	.04**	.04**	1	
4. Disruptive Behavior (DB)	-.04**	-.06**	-.44**	1

** Correlation is significant at the 0.01 level (2-tailed).

Table 10

Spearman's Rho Correlation Matrix for School Level Variables

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Academic Engaged Time School Mean (AET)	1										
2. Disruptive Behavior School Mean (DB)	-.68**	1									
3. Reading Fluency School Performance (wcpm)	.10	-.20	1								
4. Math Total Problems Correct School Performance	.30*	-.31*	.75**	1							
5. Math Digits Correct Per Minute School Performance (dcpm)	.27*	-.29*	.76**	.99**	1						
6. Off-Task Behavior School Mean (OFT)	-1.0*	.68**	-.10	-.30*	-.27*	1					
7. School-Wide Reading Performance	.12	-.18	.35**	.26*	.21	-.12	1				
8. Free and Reduced Lunch School Mean	-.43**	.30*	-.29*	-.59**	-.56**	.43**	-.24	1			
9. Special Education Status School Mean	.04	-.19	.05	.22	.25	-.04	-.04	-.18	1		
10. ELL Status School Mean	.14	.02	-.28*	-.29*	-.27*	-.14	-.47**	.35**	-.14	1	
11. Class Size	.43**	-.41**	.20	.47**	.48**	-.43**	-.08	-.48**	.18	-.05	1

** Correlation is significant $p < .001$ * Correlation is significant $p < .005$

Table 11

Assessing the Bivariate Relationships Between Different School-Level Predictors on School-Level Academic Engaged Time

School-level Predictors	Linear Regression*			Logistic Regression**		
	<i>M</i>	95% CI	<i>p</i>	<i>OR</i>	95% CI	<i>p</i>
School % of Free Lunch						
< 80% (Non-Title One School, <i>n</i> =40)	Ref			Ref		
≥ 80% (Title One School, <i>n</i> =16)	-.014	(-.040, .012)	.210	.26	(.11, .65)	.004
School % of White, Non-Hispanic						
< 30% (<i>n</i> =19)	Ref			Ref		
30% to <55% (<i>n</i> =18)	.037	(-.003, .077)	.070	2.78	(1.01, 7.62)	.048
≥ 55% (<i>n</i> =19)	.005	(-.020, .031)	.689	2.58	(.43, 15.66)	.302
School Math (Per change in 1 Digit Correct/Minute)	-.003	(-.032, .027)	.866	1.05	(.23, 4.74)	.950
School Reading (Per change in 10 Words Correct/Minute)	-.001	(-.005, .002)	.803	.90	(.75, 1.09)	.726

We fit separate models for each predictor and outcome. All models fit used GEE with exchangeable correlation to account for clustering within district and were further adjusted for region to account for regional effects.

* Linear Regression used the continuous outcome mean school-level academic engagement

** Logistic Regression used the binary outcome high mean school-level academic engagement time defined as ≥ 0.85

Table 12

Assessing the Bivariate Relationships Between Different School-Level Predictors on School-Level Disruptive Behavior

School-level Predictors	Linear Regression*			Logistic Regression**		
	<i>M</i>	95% CI	<i>p</i>	<i>OR</i>	95% CI	<i>p</i>
School % of Free Lunch						
< 80% (Non-Title One School, <i>n</i> =40)	Ref			Ref		
≥ 80% (Title One School, <i>n</i> =16)	-.006	(-.015, .003)	.180	.71	(.41, 1.21)	.200
School % of White, Non-Hispanic						
< 30% (<i>n</i> =19)	Ref			Ref		
30% to <55% (<i>n</i> =18)	-.018	(-.042, .005)	.127	1.24	(.46, 3.30)	.670
≥ 55% (<i>n</i> =19)	-.009	(-.035, .016)	.492	.57	(.18, 1.77)	.330
School Math (Per change in 1 Digit Correct/Minute)	-.008	(-.030, .014)	.467	.49	(.26, .93)	.028
School Reading (Per change in 10 Words Correct/Minute)	-.002	(-.004, .001)	.649	.84	(.80, .88)	.023

We fit separate models for each predictor and outcome. All models fit used GEE with exchangeable correlation to account for clustering within district and were further adjusted for region to account for regional effects.

* Linear Regression used the continuous outcome mean school-level academic engagement

** Logistic Regression used the binary outcome high mean school-level disruptive behavior proportion time defined as ≥ 0.10

Figure 1

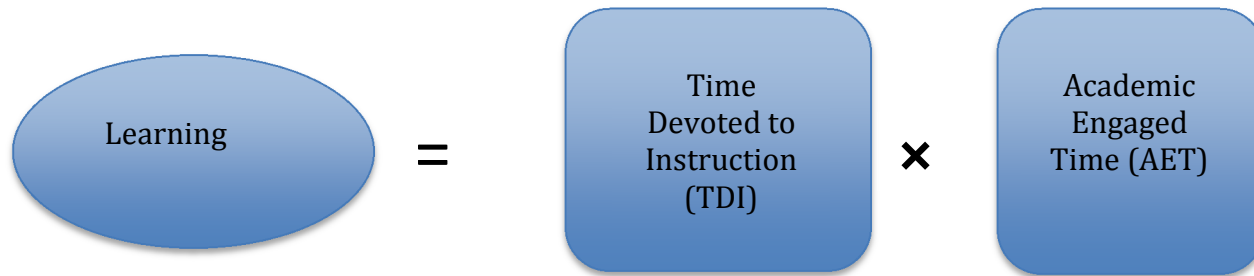


Figure 1. Time devoted to instruction (TDI) is defined as the dosage of instruction and is determined by the total amount of time instruction can be delivered. Academic Engaged Time (AET) is the degree to which a student is engaged in learning activities.

Figure 2

Partial Example of Classroom Behavioral Observation Form

School District: School Name: Teacher Name: Class Activity: Coder Name: Date: Time:	
--	--

Student Descriptor	Student Number	Minute	Interval	AE	OFFT	DB	Name of Student
	1	1	:10	1	0	0	
			:20	1	0	0	
			:30	1	0	0	
			:40	1	0	0	
			:50	1	0	0	
			:60	1	0	0	
			%	1	0	0	
	2	1	:10	1	0	0	
			:20	1	0	0	
			:30	1	0	0	
			:40	1	0	0	
			:50	1	0	0	
			:60	1	0	0	
			%	1	0	0	

Figure 3

Histogram of School Means

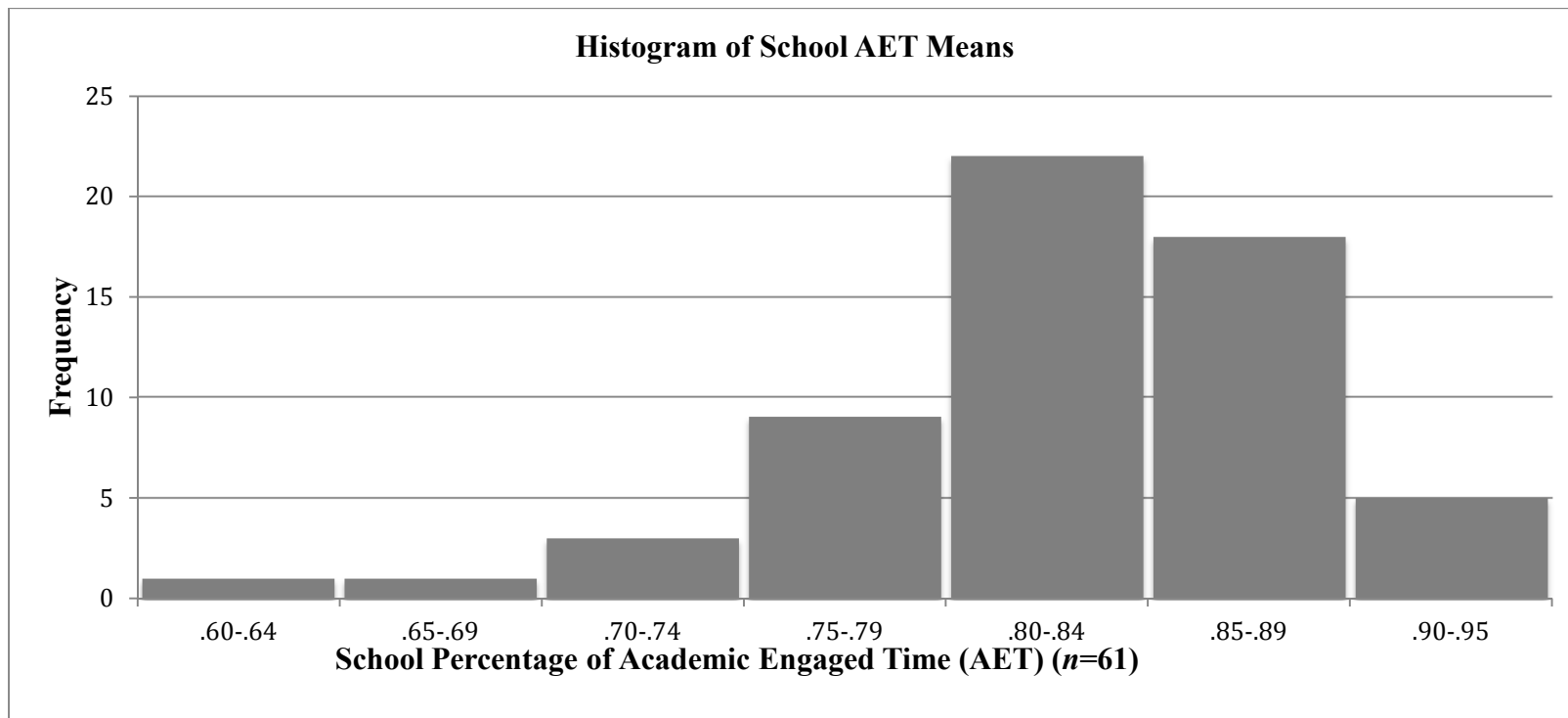


Figure 4

Histogram of Classroom Means (n=318)

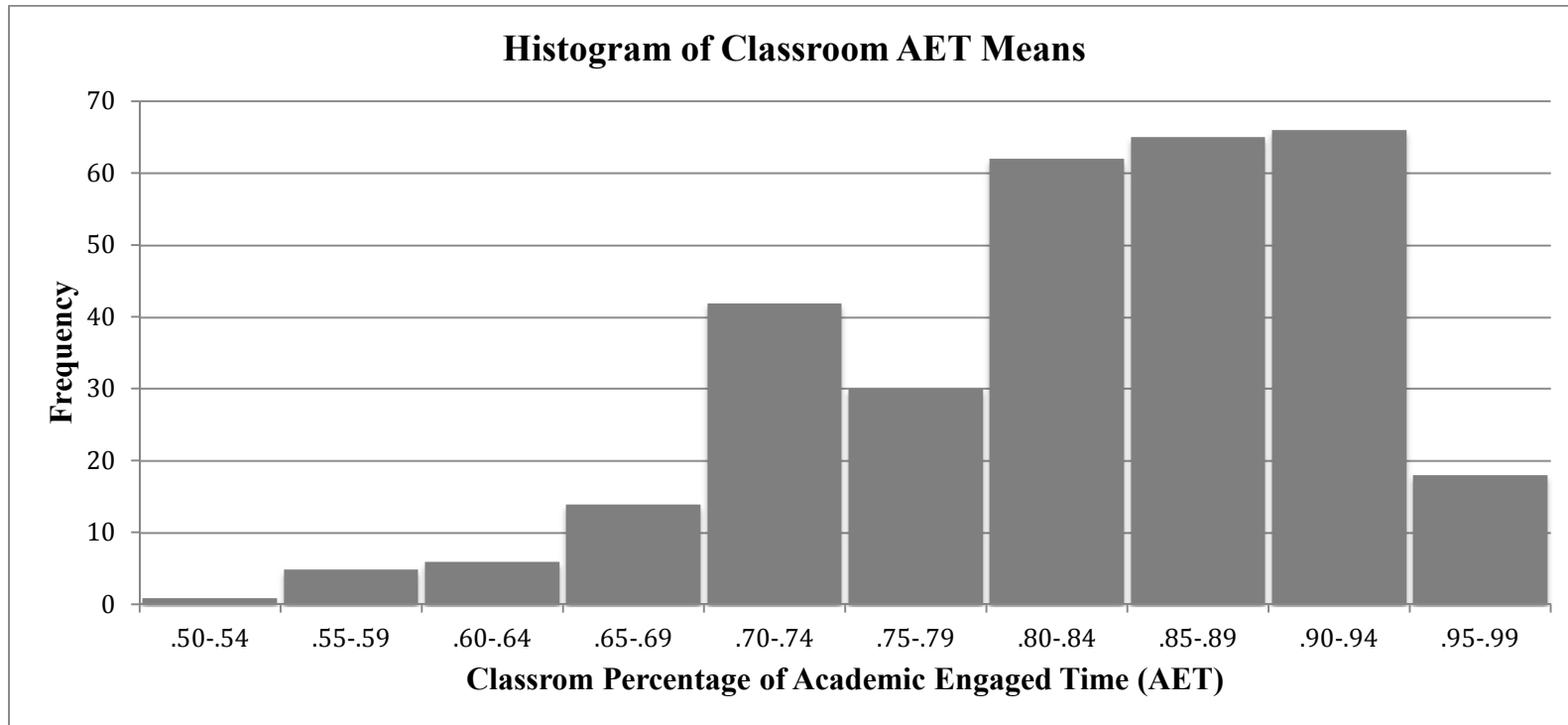


Figure 5

Federal Way School District AET Histogram

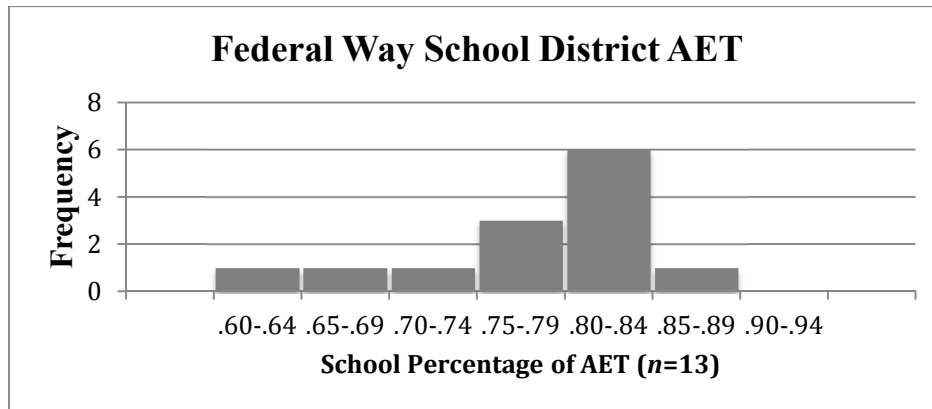


Figure 6

Lake Washington School District AET Histogram

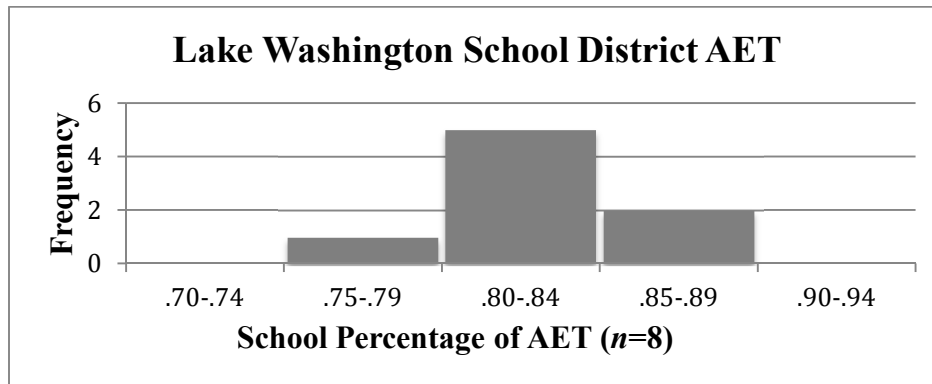


Figure 7

Mesa School District AET Histogram

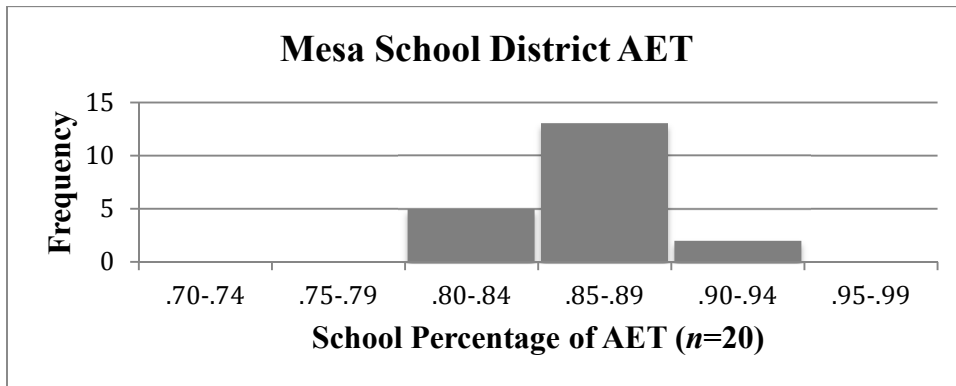


Figure 8

Mulkiiteo School District AET Histogram

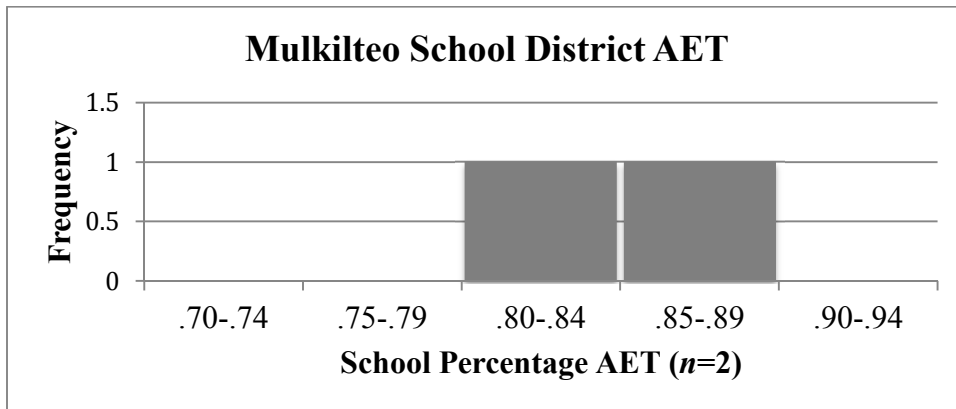


Figure 9

North Kitsap School District AET Histogram

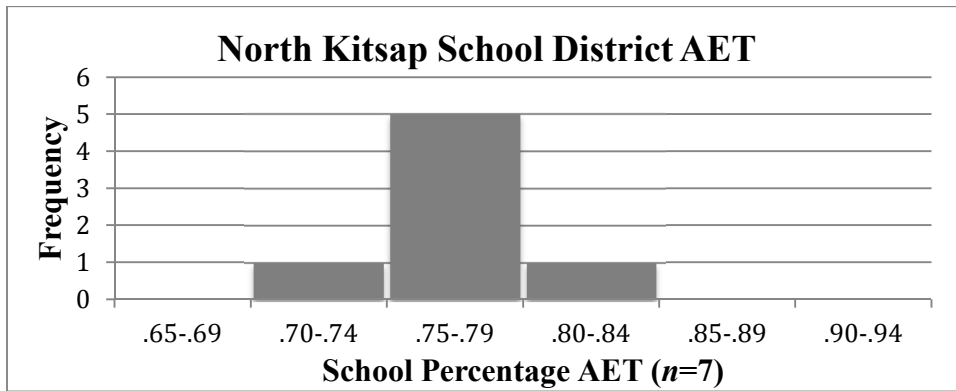


Figure 10

Renton School District AET Histogram

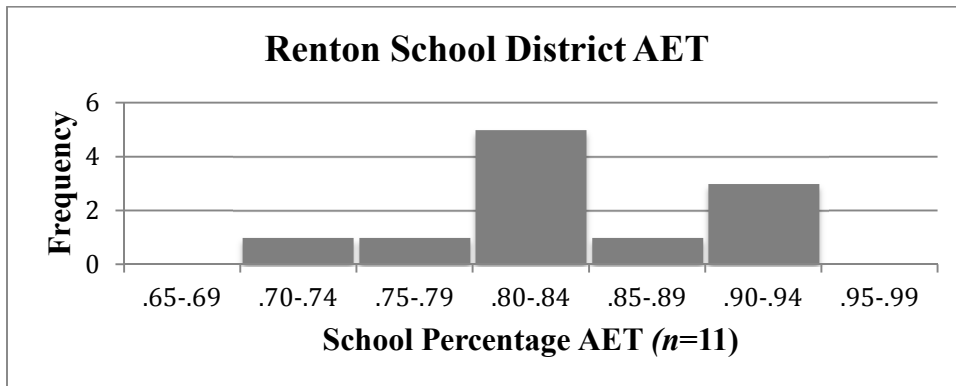


Figure 11

Regional AET Means with Confidence Interval Bars

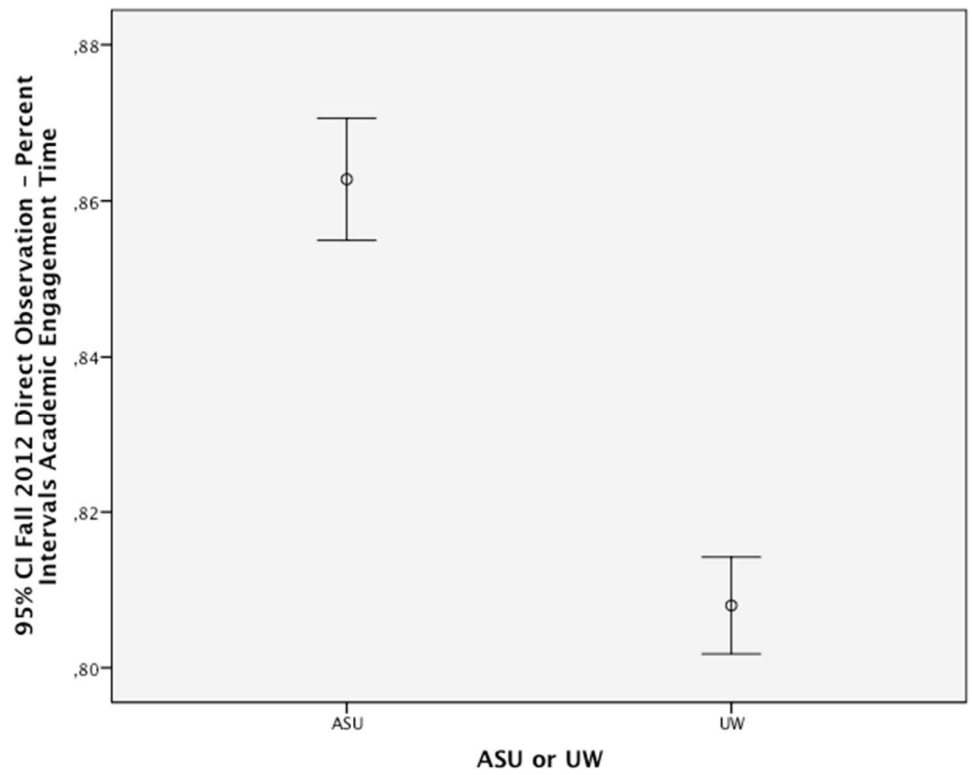


Figure 12

District AET Means with Confidence Interval Bars

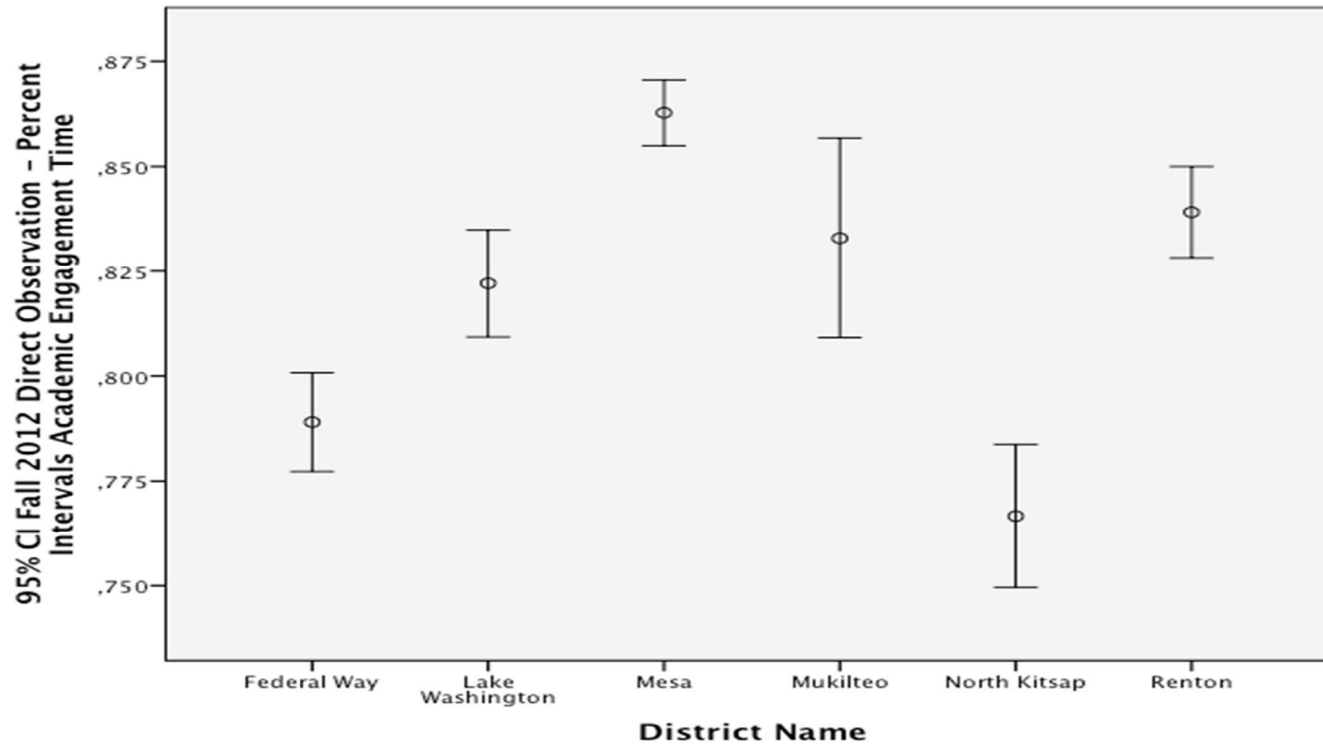


Figure 13

School AET Means with Confidence Interval Bars

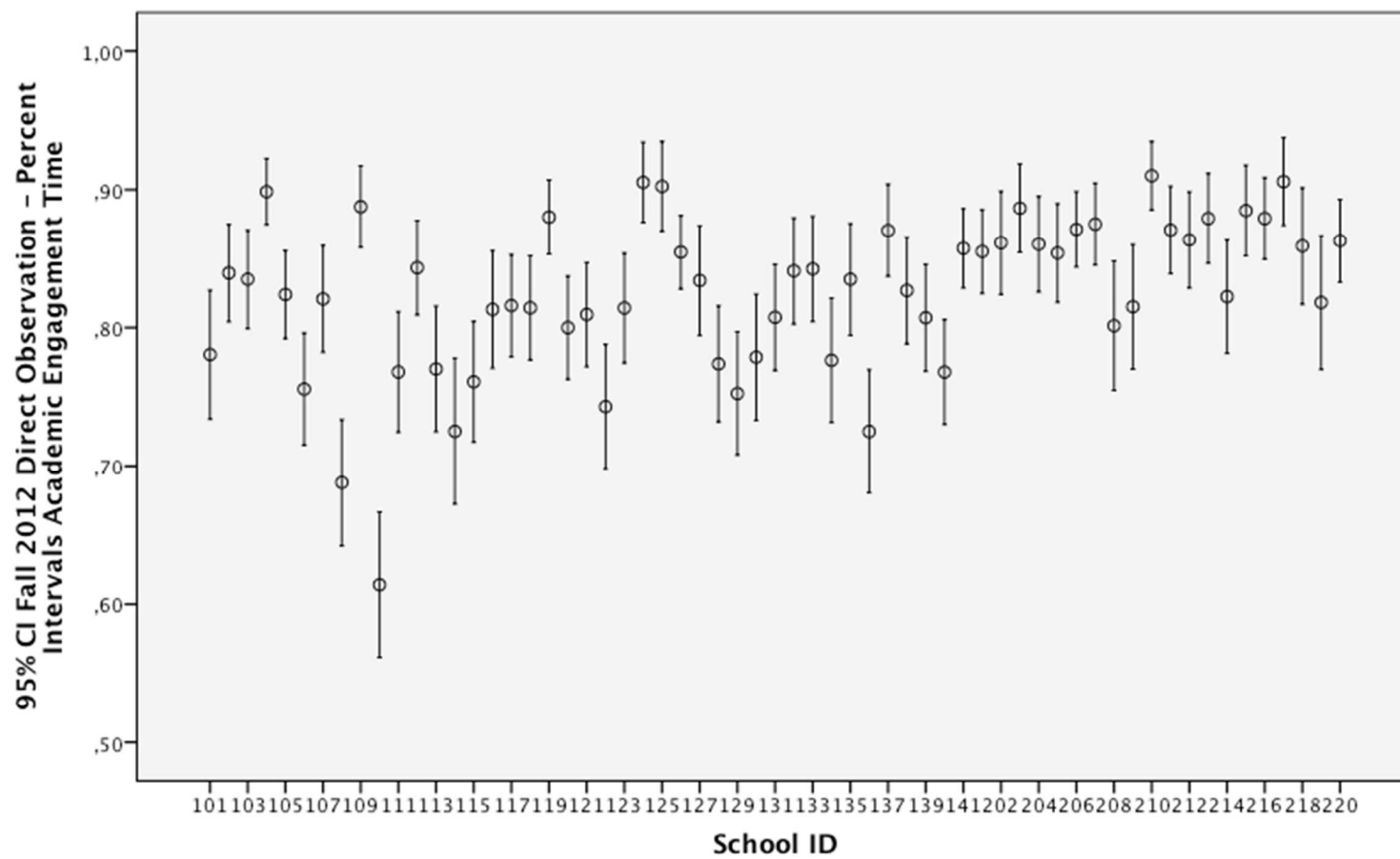


Figure 14

Histogram of School Academic Engaged Time (AET)

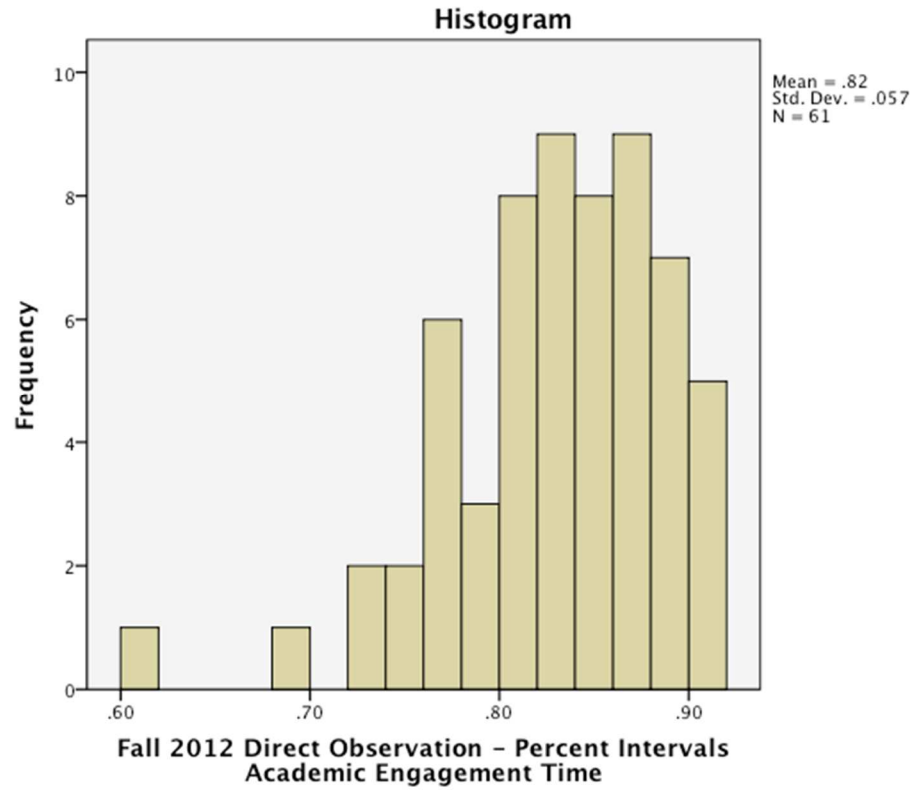


Figure 15

Histogram of School Disruptive Behavior (DB)

