LONGITUDINAL CHARACTERISTICS AND INCREMENTAL VALIDITY OF

THE STUDENT ENGAGEMENT INSTRUMENT (SEI)

by

MATTHEW DAVIS LOVELACE

(Under the Direction of Amy L. Reschly)

ABSTRACT

The two studies presented here investigated the longitudinal properties of the Student Engagement Instrument (SEI), a self-report measure of cognitive and affective engagement in school. First, change and stability were studied in SEI scores through the middle and high school years, including annual retest stability and mean level change. Next, the four-year predictive utility of the measure was assessed in terms of its ability to predict dropout and on-time graduation relative to data commonly available in school records, such as academic achievement and disciplinary data. Here, analyses of various ninth grade predictors were subjected to two-pronged tests of predictive power and yield and, ultimately, to an incremental validity analysis through a two-level multivariate logistic regression. Results indicated that stability and change in SEI scores over time fit well with expectations in terms of theory and prior empirical evidence in the engagement literature. Further, some factors on the SEI met several rigorous tests of predictive validity in relation to dropout, even when controlling for other powerful predictors of the outcome. Implications and limitations of findings are discussed, as well as future directions for research.

INDEX WORDS: Student engagement, Dropout, On-time graduation, Predictive validity, Multilevel modeling

LONGITUDINAL CHARACTERISTICS AND INCREMENTAL VALIDITY OF THE STUDENT ENGAGEMENT INSTRUMENT (SEI)

by

MATTHEW DAVIS LOVELACE

B.A., University of Tennessee, 2003

M.A., University of Georgia, 2010

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2013

LONGITUDINAL CHARACTERISTICS AND INCREMENTAL VALIDITY OF

THE STUDENT ENGAGEMENT INSTRUMENT (SEI)

by

MATTHEW DAVIS LOVELACE

Major Professor: Amy L. Reschly

Committee: Scott P. Ardoin
Jennifer H. Lindstrom
Zhenqiu (Laura) Lu

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2013

ACKNOWLEDGEMENTS

I am very grateful to those who mentored, challenged, and supported me throughout my work on this dissertation. Thank you, Amy, for your guidance, not just on this project, but throughout my entire graduate career. Also, many thanks to Colin Martin, Jim Appleton, and everyone else at the Office of Research and Evaluation in Gwinnett County Public Schools, for your assistance and support. I also wish to thank my committee members, Scott Ardoin, Jennifer Lindstrom, and Laura Lu for your valuable and constructive feedback. And, of course, thank you to my parents. I'm not sure how I would have made it through 20+ years of formal education without the safety net of your unfailing love and support throughout the years. But this dissertation is without a doubt dedicated to my wife, Mary, for the many months that you kept our lives in working order while I completed this work. By my calculations, I owe you about a lifetime of doing the dishes.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

DISSERTATION INTRODUCTION

In a 2013 provisional report, the U.S. Department of Education released data on public high school graduation rates that suggested an all-time high since 1968 for the 2009-2010 graduating class (Stillwell & Sable, 2013). This report was covered by many major news sources and may have suggested to many Americans that we were no longer a *Nation at Risk* (Gardner, 1983)—or at least that we were less at risk than we were in the early 1980s. Headline grabbing data like these, however, describe only one piece of the complicated truth about graduation and dropout rates in American schools. For one, these stories tend to focus too heavily on the *overall* rate, when there are clearly troubling trends beneath the surface, such as tremendous group differences in terms of race and income. In 2008, for instance, among 16- to 24-year-olds nationally, 9.9% of Blacks and 18.3% of Hispanics were not enrolled in school and had not earned a high school diploma or equivalent credential (Chapman, Laird, & KewalRamani, 2010), respectively nearly 2 and 4 times the 4.8% rate for Whites. Disparities like these are even wider at the college level. In 2002, enrollment rates for Hispanic and Black students were higher at 2-year and for-profit institutions than at 4-year public and private institutions (Aud et al., 2011), and the 6-year graduation rates for Hispanic and Black students at 4-year institutions—49% and 40%, respectively—were considerably lower than those of Asian/Pacific Islanders (67%) and Whites (60%). These low college completion rates may in part be due to the fact that many high school graduates are underprepared for post-secondary education. At 4-year public universities in

2008, for instance, 39% of freshmen were enrolled in remedial courses to improve basic knowledge and skills in areas like mathematics and writing (Aud et al., 2011, Fig. 22-2).

News stories about graduation rates also commonly fail to highlight the varied problems common in the data and methods used to calculate graduation and dropout rates. While one calculation method may suggest growth in graduation rates, other data sources and methods have suggested slight decline over the last 40 years—or stagnation at best (Rumberger, 2011). Regardless of the data and methods or whether or not overall growth has occurred, it is evident in all national data that each year too many students continue to drop out of school and that risk for dropout, at both secondary and post-secondary levels, is disproportionate across demographic subgroups (Rumberger, 2011).

The importance of addressing the dropout problem is underscored by the widespread and highly detrimental consequences associated with dropout, not just for the individual, but for society as well. Dropouts are more likely to commit crimes, qualify for and receive welfare, and have poorer health outcomes, as well as being less likely to vote and engage in other civic activities (Rumberger, 2011). These individual-level outcomes can lead to rippling effects that have an impact on surrounding lives, neighborhoods, and communities—reducing the quality of life for victims of crime, increasing the cost and availability of health services, and decreasing the desirability and median property value in a neighborhood. Economists have further estimated significant effects of dropout on the broader economy. Rouse (2005) estimated that, in addition to costs related to issues like increased crime and poorer health, a single cohort of dropouts could account for over $200 billion in foregone income and tax revenues in their working lifetime.

Dropout and Student Engagement

The path to a successful graduation is a complex process involving a dynamic interplay between an innumerable set of risk and protective factors. One variable, however, that has caught the attention of many researchers and policy-makers is student engagement at school and with learning. Over the past two decades, engagement has become increasingly viewed as a necessary centerpiece in efforts to address important educational problems, due in part to its alterable nature (Lehr, Johnson, Bremer, Cosio, & Thompson, 2004) and to findings of its associations with academic achievement, lower-risk health behaviors, well-being, and other long-term outcomes, including work success and school completion (Christenson, Reschly, & Wylie, 2012). For these reasons, engagement has been featured prominently within dropout prevention programs (e.g., Check & Connect; see Christenson et al., 2008) and comprehensive school reform efforts (e.g., Talent Development High Schools; see Legters, McPartland, & Balfanz, 2004) designed to address the low achievement, alienation, and high dropout rates that many students experience, particularly those from disadvantaged backgrounds.

Early studies of engagement primarily utilized behavioral indicators (e.g., time on task, being prepared for class) to measure the construct, but arguments to operationalize and measure engagement multi-dimensionally (Appleton, Christenson, Kim, & Reschly, 2006; Appleton, Christenson, & Furlong, 2008; Fredricks, Blumenfeld, & Paris, 2004) have led to a shift toward including a variety of indicators in studies of the construct, including cognitive and affective engagement. It is challenging to study these internal aspects of engagement, because reliable and valid information about such latent characteristics are less available than behavioral data. Indicators such as frequency of disciplinary referrals and number of extra-curricular hours are readily collected by schools as a part of standard record keeping, variables which provide an efficient and accessible view into many important behaviors of students. But because thoughts and feelings about peers, teachers, and schools are not directly observable, it can be difficult to glean such information from school records for the vast majority of students in a systematic and reliable way. One way that schools may attempt to gather data on students' cognitive and affective engagement is through the use of psychometric instruments.

Most of the behavioral data gathered for school records that is relevant to student engagement—such as data on absences and office disciplinary referrals—would best be described as tapping into how *disengaged* students are. Data on which students are disengaged is valuable, but it says little about the students who are not behaving badly. In other words, if behavioral engagement in school were imagined to exist along a continuum that spanned from *highly disengaged* to *highly engaged*, a student who attends regularly, is not late to class, and has never been sent to the office could, at best, be considered *not highly disengaged* based on this information alone. But this is not a satisfying answer to a question like, "How engaged are most students at this school?" This practical issue further supports an argument for systematically collecting data on cognitive and affective engagement in schools, because rather than having to rely solely on crude dichotomies like behaviorally disengaged/not-disengaged, a more efficient way to meaningfully further differentiate students could be through the use of short questionnaires used to measure thoughts and feelings related to engagement.

Several engagement measures have been developed to fit within a multi-dimensional model (Fredricks et al., 2011), one of which is the Student Engagement Instrument (SEI; Appleton et al., 2006). Empirical evidence from a variety of studies supports the use of the SEI in measuring cognitive and affective student engagement. Favorable results have been found regarding internal consistency (Appleton et al., 2006), meaningful factor structure (Appleton et al., 2006; Betts, Appleton, Reschly, Christenson, & Huebner, 2010), measurement invariance from grades six through twelve (Betts et al., 2010), and association with external measures of academic functioning and school behavior (Lovelace, 2010, 2011). Some aspects of the SEI's reliability and validity, however, have not yet been fully investigated.

One under-researched area of the SEI is its performance across time. Now that the SEI has been used in its present form in schools for several years, it is possible to investigate the stability and change of students' SEI scores over time, as well as how such time-related aspects of the SEI fit with theory and prior empirical evidence on engagement. Furthermore, if the SEI is shown to perform adequately over time, a second question that can begin to be addressed is whether predictive information is gained by measuring cognitive and affective engagement with the SEI, beyond what is explained by other data already available in school records. The following two studies addressed these broad questions, investigating (1) the longitudinal nature of scores on the SEI and (2) the incremental validity of the measure.

In Study 1, presented in Chapter 2, aspects of stability and change in SEI scores over time were examined. Questions were addressed related to the retest stability of scores from one year to the next, what SEI score trajectories looked like for the average student, and whether individual-level patterns of change were suggested by patterns over time. In the literature review, findings from previous validation studies of the SEI are summarized in further detail, in addition to a focus on what is known about stability and change in engagement through adolescence. Through an accelerated longitudinal design, Study 1 was conducted using SEI responses from a variety of overlapping cohorts, collectively representing the responses of over 40,000 middle- and high-school students, all of whom completed the SEI each semester from Fall 2008 to Spring

2011. Estimates of annual retest stability are presented, as well as findings from an investigation into group-level and individual-level change over time.

In Chapter 3, Study 2 is presented, in which SEI scores were used as explanatory variables to investigate the incremental validity of the instrument over data commonly available in school records. Previous studies have found associations between SEI scores and educational outcomes (Appleton et al., 2006; Carter, Reschly, Lovelace, Appleton, & Thompson, 2012; Lovelace, 2010, 2011), but an important practical question remained. Of the variance explained in relevant outcomes like dropout and graduation, is some of it unique to SEI scores, or can it be completely explained by data that are immediately accessible to schools? In Study 2, the primary focus of the literature review is on the associations between various indicators of engagement and measures of school performance and completion, as well as on the importance of establishing incremental validity for measures used in applied settings. Next, analyses on the predictive efficiency (positive predictive value, sensitivity) of ninth grade indicators are presented, in which the utility of a variety of individual indicators was assessed through a stringent two-pronged test. Finally, analyses on 4 years of data are presented, in which the incremental influence of SEI factors on the odds of dropping out or graduating on time were estimated via multilevel logistic regression modeling while controlling for known powerful predictors already available in school data, such as attendance, discipline data, and standardized achievement performance.

CHAPTER 2

STABILITY AND CHANGE IN A MEASURE OF COGNITIVE AND AFFECTIVE STUDENT ENGAGEMENT[1]

[1]Lovelace, M.D., A.L. Reschly, and J.J. Appleton. To be submitted to *Journal of Adolescence*.

Introduction

A growing number of educators, practitioners, policymakers, and researchers are focusing on engagement as a necessary factor in efforts to address and understand the high rates of underachievement, disengagement, and dropout facing high schools, as well as in efforts to increase post-secondary enrollment and persistence (Appleton, Christenson, & Furlong, 2008; Christenson, Reschly, & Wylie, 2012; Fredricks, Blumenfeld, & Paris, 2004; Fredricks et al., 2011; National Research Council and the Institute of Medicine, 2004). Research on student engagement has been active for over 25 years, and although this work has led to general agreement among researchers on various aspects of engagement definitions, theory, and research, there are questions and issues that remain to be addressed (Reschly & Christenson, 2012). Debates persist, for example, about how to best conceptualize student engagement, how it differs from motivation, and how to measure student engagement most effectively.

Most of the international research community that studies student engagement conceptualizes the construct as multi-dimensional (Reschly & Christenson, 2012), typically involving some aspects of behavior, emotion, and cognition (Fredricks et al., 2004; Reschly & Christenson, 2012). Appleton, Christenson, and others use these dimensions, but further differentiate behavioral engagement into two sub-types: behavioral and academic engagement (Appleton, Christenson, Kim & Reschly, 2006; Christenson et al, 2008). Across scholars, there is still a great deal of variation regarding which subtypes or dimensions are measured and included in analyses (Janosz, 2012). For example, some studies emphasize behavioral aspects only, while others may emphasize both behavioral and cognitive indicators, in both cases ignoring or not measuring affective aspects of engagement. Further, even within the general agreement of multidimensionality, researchers define the subtypes differently (Reschly & Christenson, 2012).

For example, a student's perceived relevance of school is viewed as affective engagement by Finn (2006), whereas this same quality is considered cognitive engagement by Appleton et al. (2006). Many studies have indicated that student engagement is a substantial predictor of persistence in school and high school dropout (Rumberger & Lim, 2008), but because of theoretical and research differences like these, the contributions made by different engagement dimensions in predicting student outcomes like dropout and persistence are less known (Janosz, 2012).

Amid the increased attention being paid to student engagement, a variety of subscales and instruments have been developed to measure student engagement (Fredricks et al., 2011). One of these tools is the Student Engagement Instrument (SEI), designed to measure cognitive and affective engagement with school (Appleton et al., 2006). The SEI was designed to measure students' internal experiences of engagement, data which could be used to complement information on observable indicators of academic engagement (e.g., homework completion) and behavioral engagement (e.g., attendance) that are regularly available in school records (Appleton et al., 2006; Fredricks et al., 2011). If the longitudinal reliability and validity of the SEI could be well established, when coupled with school data, the SEI would have the potential to advance our understanding of student engagement by allowing researchers to study and better understand the degree to which indicators of cognitive, affective, academic, and behavioral engagement independently and interactively affect student outcomes.

Conceptual Basis of the SEI

The SEI is a self-report survey designed to measure student self-perceptions of engagement with school. It taps into facets of cognitive and affective engagement through a five-factor model (Appleton et al., 2006), and according to the developers of the SEI (Appleton et al., 2006, 2008) it was designed within a theoretical framework based on the work of school completion researchers (e.g., Finn, 1989) and on the implementation and evaluation of the Check & Connect school intervention model (see Christenson et al., 2008). Within the SEI's conceptual framework, student engagement is viewed as a construct consisting of four broad dimensions of engagement: cognitive, affective, behavioral, and academic (Appleton et al., 2006, 2008). Furthermore, the Appleton model suggests that a dynamic relationship exists between context, engagement, and student outcomes—an ecological perspective that considers the goodness-of-fit between student, learning environment, and various influential factors (Appleton et al., 2006, 2008; Christenson et al., 2008).

Findings from Prior Studies of the SEI

Appleton et al. (2006) developed the initial items on the SEI after a review of relevant literature, and later refined SEI items through pilot studies on diverse focus groups. The SEI's factor structure was discovered through exploratory factor analysis (EFA) of SEI responses of one-half of a sample of 1,931 ninth graders. Decisions regarding the number of factors to retain were initially conducted through a combination of scree plot analysis and the K1 procedure, which suggested retaining four to six factors. Further decisions about factor structure were made through an iterative review of four-, five-, and six-factor structures with EFAs until all items loaded at .40 or higher. These models were subsequently subjected to confirmatory factor analysis (CFA) using the other half of the sample, which suggested that the five- and six-factor models fit the data better than the four-factor model. Internal consistency estimates for the SEI ranged from α = .72 to .88. Additionally, in an analysis of convergent and divergent validity, Appleton et al. reported moderate correlations in expected directions between SEI factors. Small correlations were found, generally in expected directions, between SEI factors and several educational outcomes (i.e., GPA, standardized achievement scores, and suspensions).

Since the initial pilot study, follow-up validation studies have been conducted. First, Betts, Appleton, Reschly, Christenson, and Huebner (2010) conducted an analysis of the SEI's reliability and factorial invariance across multiple grade levels on a sample of 2,416 sixth to twelfth grade students drawn from school districts in the rural Southeastern and Upper Midwestern United States. Analysis of the SEI's factor structure revealed that items and factors conformed to results of the original SEI study. Using multiple statistical criteria to judge the invariance of the model across grade levels, a five-factor model with a consistent structure was found to fit well across students in grades 6 to 12.

Reschly, Betts, and Appleton (2013)—in a study of two engagement measures, the SEI and the Motivation and Engagement Scale (MES; Martin, 2003)—further evaluated the model fit and convergent and divergent validity of the SEI with a sample of adolescents (N = 227). CFA indicated acceptable fit of the SEI, while MES fit was poor. Correlations between SEI and MES factors, as well as correlations between SEI factors and external indicators of academic functioning and behavior in school, supported earlier findings of the convergent and divergent validity of the SEI.

Objectives of the Present Study

Longitudinal characteristics of the SEI were assessed in the present study. The broad purpose was to investigate the stability of the instrument across repeated administrations with the same students, as well as normative patterns of change in adolescents' SEI scores over the course of middle and high school. Taking a comprehensive view of stability and change, the objectives were to (a) estimate the annual retest stability of the SEI, (b) describe—at the group level—the extent to which SEI scores change as a function of time; and (c) investigate whether individual-level patterns of change in SEI scores occur for many students. Based on these broad objectives, three essential research questions were addressed:

(1) What is the annual retest stability of cognitive and affective engagement as measured by the SEI?

(2) As theory and relevant research would suggest, do SEI scores decrease as a function of time through the adolescent years for most students?

(3) Considering the annual stability of the instrument, is there also evidence of individual-level change in SEI scores?

The first research question sought to examine the stability of the SEI from year to year. As noted above, Appleton et al. (2006) and Betts et al. (2010) reported good internal consistency estimates for the SEI with a variety of samples. Furthermore, evidence from cross-sectional data suggested that the factor structure of the SEI is invariant across groups of students from grades 6 to 12 (Betts et al., 2010). Retest stability, however, has not been estimated for the SEI. Estimating stability is important to understanding the psychometric value of an instrument (Crocker & Algina, 2008), and with an emerging construct like engagement it can also provide insight into the very nature of what it measures (Asendorpf, 1992).

The second research question addresses normative change in SEI scores through middle and high school. Normative change, also known as absolute change or mean-level change, can be used to reveal patterns of development that apply to most students (Roberts, Walton, &

Viechtbauer, 2006). Here, normative change refers to mean-level change in SEI scores in middle and high school students over time. Applied research with cross-sectional data suggests gradual decreases over time in SEI scores (Appleton, 2012), but no accelerated or true longitudinal designs have explored this question.

The third research question refers to seeking evidence of change beyond the normative trend. There are few longitudinal studies on student engagement, but evidence to date has indicated that, while most students follow common pathways of engagement, sub-trajectories may also be evident (Janosz, Archambault, Morizot, & Pagani, 2008; Wylie & Hodgen, 2012). Janosz et al. (2008), for example, found evidence of several non-normative trajectories of engagement, and some of these pathways were substantial predictors of dropout.

Hypotheses

Stability. Evidence of moderate levels of annual retest stability was expected for all factor scores on the SEI. This hypothesis was informed by Conley's (1984) study of the *hierarchy of consistency*, a theory of predictable differences in the longitudinal stability of various classes of psychological constructs according to their respective position on a state-trait continuum. Conley found that constructs theorized to be the most trait-like showed the most longitudinal stability (e.g., intelligence), followed by personality constructs (e.g., extraversion), and then by relatively stable but much more state-like constructs, like happiness, self-esteem, and life satisfaction. Based on the hierarchy of consistency and Conley's findings, cognitive and affective engagement—conceptualized as context dependent and therefore more state-like than trait-like in nature—were expected to be moderately stable from year to year.

Expectations for the observed retest stability of SEI factor scores (see Table 2.1) were derived from findings presented in Conley's (1984) paper and through a formula he presented:

$$C = Rs^n$$

where C is the observed retest stability coefficient, R is the internal consistency of the instrument, s is the annual stability of the construct, and n is the interval in years over which the coefficient is calculated. In this formula, expected observed values for the stability of the measure (C) are calculated in a way that separates the effects of measurement reliability (R) from the true longitudinal stability of a construct (s). The value .70 was considered an appropriate approximation for the annual consistency of the engagement constructs based on the annual stability of the state-like constructs in Conley's study most similar to engagement (i.e., self-esteem, happiness), which were based on data from studies involving adolescent-aged participants. Estimates for R were generated using the means of internal consistency coefficients reported across two studies of the SEI (i.e., Appleton et al., 2006 and Betts et al., 2010), each of which used a different approach than the other to estimate the period-free reliability of the instrument.

Table 2.1

Hypothesized annual retest stability for each SEI factor

SEI factor	R	s	C (annual)
Teacher-Student Relationships (TSR)	.81	.70	.56
Family Support for Learning (FSL)	.78	.70	.54
Peer Support for Learning (PSL)	.79	.70	.55
Future Goals & Aspirations (FGA)	.79	.70	.55
Control & Relevance of Work (CRW)	.75	.70	.52

Note. C = observed retest stability coefficient, R = internal consistency of the measurement instrument, and s = annual stability of the construct.

Individual-level change. A stability estimate close to +1.0 or -1.0 would suggest very little between-student differences in within-student change over time (Asendorpf, 1992). But if, as expected, the annual retest stability of SEI factor scores are no more than moderately stable, this would suggest that a good number of students' SEI scores may vary considerably over time. This kind of *differential stability* (Asendorpf, 1992) should be expected with a context dependent variable like engagement. It is reasonable to assume that a subset of students experience wide differences in the degree to which environmental variables that influence engagement (e.g., family, school) remain stable, and perhaps even in the degree of control these students have over such variables. Variation in contexts like these should result in noticeable differences in the direction and magnitude of individual-level change.

It is also important to note that change trends at the group level can conceal mutually canceling differences on an individual level (Blonigen, Hicks, Krueger, Patrick, & Iacono, 2006). In the mean-level change analysis, if SEI scores gradually decrease over time as expected, it would still be important to consider individual-level change to get a fuller picture of longitudinal trends in cognitive and affective engagement. In studies by Janosz et al. (2008) and Wylie and Hodgen (2012), the majority of students showed moderate levels of engagement with a slight trend downward over time. But smaller percentages of the sample followed other engagement paths. Some students in each of these studies were characterized by consistently high, if gradually decreasing, levels of engagement, while others showed much less stability, with scores either trending from low to moderate levels or from moderate to low levels or simply decreasing rapidly over time. Trends in SEI data were expected to fit with these findings; that is, although most students were hypothesized to show a normative engagement trajectory, reliable evidence of substantial change over time was expected for a subset of students.

Method

Sample and Design

Data for the present study were provided by the Office of Research and Evaluation at Gwinnett County Public Schools (GCPS), a district located in the metro Atlanta area. With over 160,000 students (Georgia Department of Education, 2013), GCPS is the largest school district in Georgia and ranks in the top 15 largest districts in the U.S. in terms of enrollment. The SEI has been used by GCPS since 2007 as part of a district-wide student advisement program geared toward enhancing student engagement (Appleton, 2012), and it is administered each fall and spring to all students, making GCPS a particularly suitable location for studying the SEI's longitudinal characteristics. The de-identified data provided by GCPS for this study contained SEI responses from all middle and high school students who completed the survey between 2008 and 2011, as well as data about gender, ethnicity, and grade level for all students enrolled each year. This study received the approval of the district's and university's institutional review boards.

An accelerated longitudinal design (Tonry, Ohlin, & Farrington, 1991; see Duncan, Duncan, & Hops, 1996 for a good example) was used to address the various objectives of this study, combining information from five separate but overlapping grade-level cohorts (N = 41,989) in which each were followed for three years (see Table 2.2). The obvious disadvantage of this design was its required assumption that linking overlapping longitudinal cohorts together would provide information about stability and trajectories that would closely resemble what would be found if a full longitudinal design were possible. Empirical evidence, however, provides support for this assumption in such designs. In a methods study comparing a single longitudinal sample design to an accelerated design, Duncan, Duncan, and Hops (1996) found no significant differences between these longitudinal methods in their representation of initial status and growth, nor differences in effects for covariates. Further, as Miyazaki and Raudenbush (2000) point out, the accelerated design offers considerable advantages when costs of data collection or risk of attrition are high. In the case of the present study, analyzing data representing three years of student responses rather than seven was not only less costly in time and resources but also reduced the chances of attrition that would have resulted from students leaving the district over time. Table 2.2 provides descriptive statistics about each of the five cohorts observed from Fall 2008 through Spring 2011.

Measures and Variables

The primary variables of interest in this study were students' factor scores on the SEI. The SEI is a self-report survey designed to measure student perceptions of their cognitive and affective engagement in school (Appleton et al., 2006). The SEI measures five facets of cognitive and affective student engagement, namely: Teacher-Student Relationships (TSR), Control and Relevance of Work (CRW), Peer Support for Learning (PSL), Future Goals and Aspirations (FGA), and Family Support for Learning (FSL). TSR, PSL, and FSL were formulated to measure aspects of affective engagement, whereas CRW and FGA were formulated to measure cognitive engagement. The SEI uses a five-point scale ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree*, and it was coded so that higher scores signify higher levels of engagement. Factor scores were computed as the mean of item responses within a factor, as is customary for the instrument (Appleton et al., 2006; Fredricks et al., 2011). Two items representing a sixth factor called Extrinsic Motivation have been included in other work (e.g., Appleton et al., 2006). Betts et al. (2010), however, did not include the Extrinsic Motivation factor in analyses, noting that there were not enough items loading onto the factor.

Table 2.2

Accelerated longitudinal design and descriptive statistics for gender and ethnicity

	Grade						
Cohort	6	7	8	9	10	11	12
Cohort 1							
n	8,908	8,330	7,929	--	--	--	
P female	.50	.51	.51	--	--	--	
P white	.36	.36	.36	--	--	--	
Cohort 2							
n	--	8,803	8,339	7,812	--	--	--
P female	--	.50	.50	.51	--	--	--
P white	--	.36	.36	.37	--	--	--
Cohort 3							
n	--	--	8,644	8,117	6,997	--	--
P female	--	--	.52	.52	.53	--	--
P white	--	--	.36	.36	.39	--	--
Cohort 4							
n	--	--	--	8,450	7,071	6,195	--
P female	--	--	--	.51	.52	.53	--
P white	--	--	--	.38	.41	.44	--
Cohort 5							
n	--	--	--	--	7,184	6,095	6,497
P female	--	--	--	--	.54	.54	.54
P white	--	--	--	--	.42	.44	.42

Note. Each cohort observed for three years from fall 2008 through spring 2011. *P* represents the proportion of students in the sample with the given characteristic. Dashes indicate that data were not available at that time-point for the cohort due to the accelerated longitudinal nature of the study.

For these same reasons, only the 5-Factor Model was included in the present study. The SEI was administered according to standardized procedures (see Appleton, 2012), including reading items aloud to students to limit effects of varying student reading levels. The primary explanatory variable in this study was time. In analyses of mean- and individual-level change, the explanatory effects of gender were also assessed. Time in the growth model analysis was measured in years between test administrations, that is, from one fall semester to the next.

Analyses

Stability. For analyses of stability, in order to control for developmental phenomena that may affect all students—such as the expected steady decrease in average engagement through adolescence or perhaps an increase in variance with increasing age—rank-order stability coefficients were computed, i.e., correlations between students' relative standing to one another based on SEI scores from one administration to the next (Asendorpf, 1992; Roberts & DelVecchio, 2000). Exploratory plots of SEI factors (see Figure 2.1) showed that the distribution of SEI factor scores were often skewed conditional on grade level. Rank-order Spearman's rho coefficients, robust against skewness or outlier problems, were therefore computed to assess year-to-year stability. To model the uncertainty in these estimates, 95% confidence intervals (CIs) were calculated using methods described by Caruso and Cliff (1997).

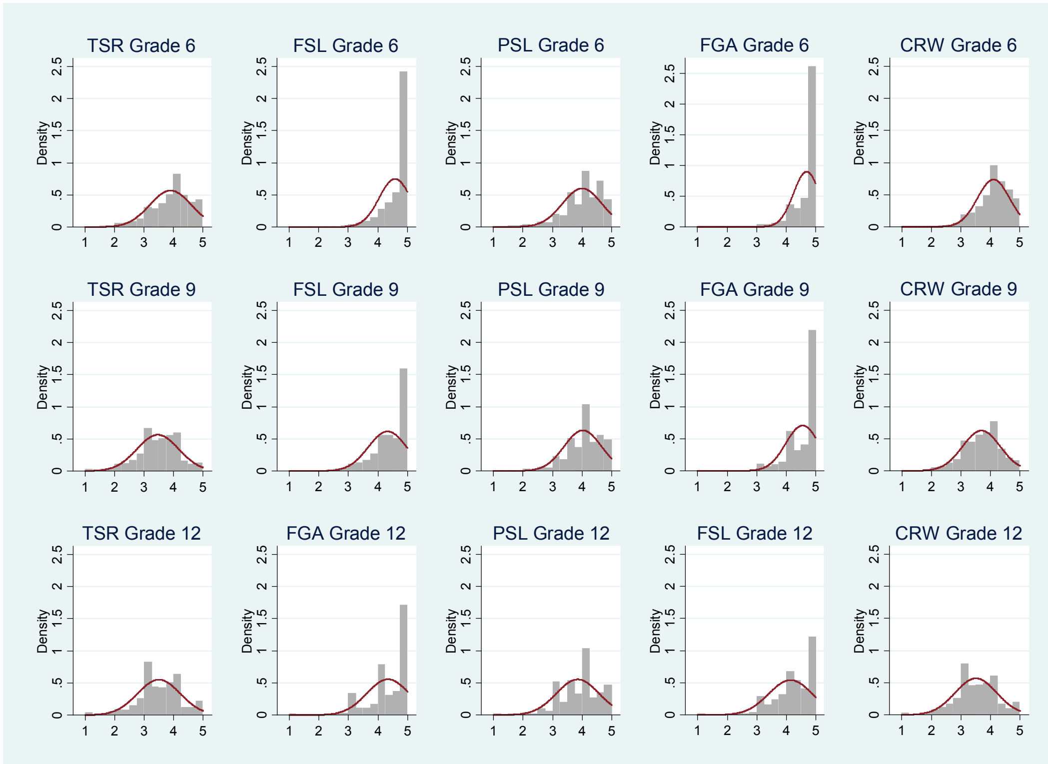

Figure 2.1: SEI factor score distributions in grades 6, 9, and 12. Normal density plots (red line) based on *M* and *SD* of the sample data.

Mean-level change. To investigate the effects of time on SEI factor scores, means over time were plotted to inform model building (e.g., whether change appeared linear or curvilinear) and to examine whether overlapping cohort trajectories appeared to link together well. In addition, Cohen's *d* statistics (Cohen, 1992) were calculated to describe, in standard deviation units, the direction and magnitude of change in the population means between the highest and lowest point of each trajectory. Next, inferential analyses involving multi-level/hierarchical techniques (Gelman & Hill, 2007; Rabe-Hesketh & Skrondal, 2012; Raudenbush & Bryk, 2002) were used to model the effect of time on SEI total scores (i.e., the mean of combined subscale scores), and to assess whether trajectories varied by gender. The decision was made to model

SEI total scores rather than each subscale separately because these distributions were much more normally distributed (see Figure 2.2). A variety of model types were fit to the data, including a basic quadratic model, a cross-level interaction model, and models with complex level-one and level-two variation—with the goal of finding an approach that best modeled and explained the data. Stata software (StataCorp, 2011) was used to estimate these growth models. To test the hypothesis of a gradual decrease in engagement for most students, slope estimates for time were examined, and their 95% CIs were used to assess statistical significance.

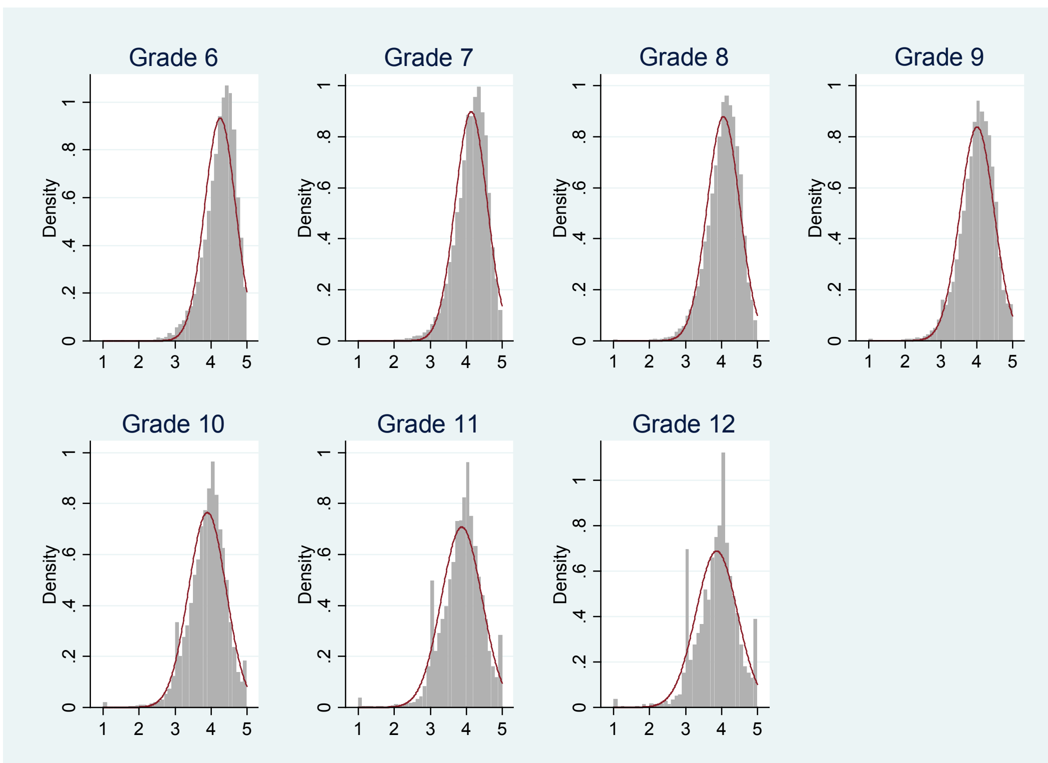

Figure 2.2: SEI total score distributions from grade 6 to 12. These histograms with normal density curves appeared to be more normally distributed overall than the SEI subscale distributions shown in Figure 2.1. Less normality is seen in later grades, suggesting that either the nature of the population, the construct, or the validity of the instrument began changing in grade 11. Based on demographic changes shown in Table 2.2, population change seemed to be the most salient possibility.

Individual-level change. To address the third research question, students were classified as having decreased, increased, or stayed the same on each SEI factor score over 2 year intervals using the Reliable Change Index (RCI; Jacobson & Truax, 1991), which describes the probability of observing a difference score equal to or greater than the one observed (Blonigen et al., 2006; Jacobson & Truax, 1991). By accounting for measurement error, the RCI was particularly useful in separating true change in SEI scores from change due to moderate levels of retest stability (Robins, Fraley, Roberts, & Trzesniewski, 2001). The RCI for each student was computed by dividing his or her change score from one point in time to another ($x_2 - x_1$) by the standard error of the difference (S_{diff}) between the two scores (Jacobson & Truax, 1991). The S_{diff} describes the distribution of difference scores that would be expected if no change had actually occurred, and it was computed in this study using the standard error of measurement (S_E) for each SEI factor score at Grades x_1 and x_2 through a formula described by Blonigen et al. (2006): $S_{diff} = \sqrt{(S_{E,x2})^2 + (S_{E,x2})^2}$.

Results

Annual retest stability

Figure 2.3 shows the observed retest stability *C* for each SEI factor score plotted against the expected retest values (dashed lines). Some notable deviations from the expected stability estimate were observed. FGA was less stable than expected from grade 6 to 7, and CRW overall was more stable than expected. In general, however, annual stability estimates fell in the moderate range, most were close to the hypothesized value, and there was excellent agreement across overlapping cohort estimates. For the most part, no substantial shifts in stability across time were observed, with stability estimates tending to fall between .50 and .60 across the middle and high school years.

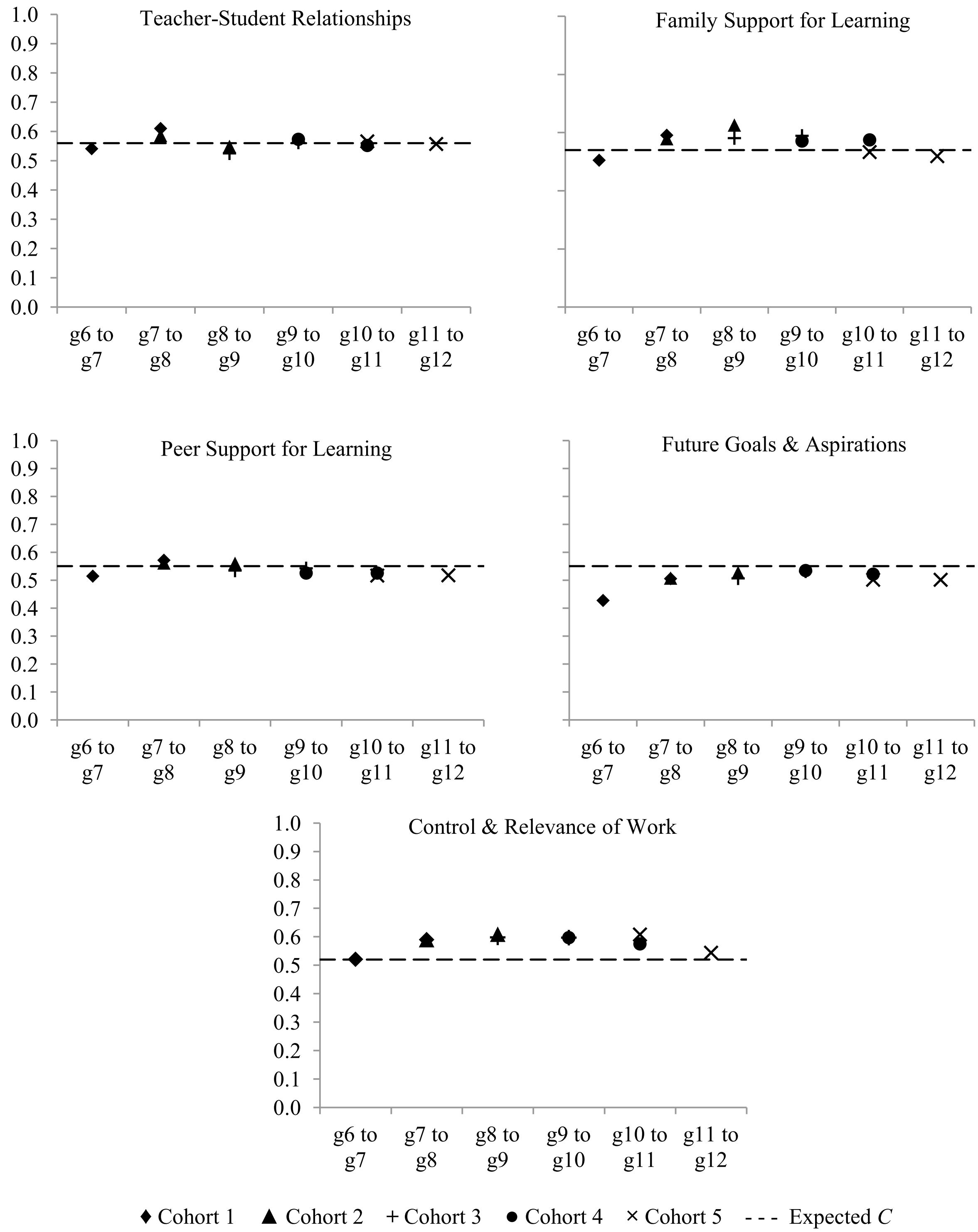

Figure 2.3: Annual retest stability of each SEI factor. Annual retest stability (represented by Spearman's rho) are plotted along the y-axis and are grouped by grade comparisons along the x-axis. Dashed lines represent the expected annual observed retest coefficient *C* based on internal consistency and hypothesized construct stability (Conley, 1984).
g = grade (as in *g8* for *grade 8*)

Group-Level Change through Middle and High School

Semester-to-semester mean-level change. Figure 2.4 displays the means for each grade by semester and cohort. As anticipated, but to varying degrees, all factor scores appeared to have gradually decreased over time. TSR and CRW appeared to change the most over time and these trends appeared curvilinear, a pattern that is common in time effects. Time trends for FSL, PSL, and FGA appeared more linear. Another interesting trend across factors was the semester-to-semester rise and fall in slopes, most notable in TSR and CRW.

Growth models for SEI total scores. Growth curve modeling, a special case of multilevel modeling in which the coefficient of time is allowed to vary randomly between students (Rabe-Hesketh & Skrondal, 2012), was used to further investigate the trajectories of student's SEI scores as they progressed through middle and high school. Here, the Level 1 units were occasions (i) structured in an accelerated longitudinal format within individual students (j). The primary goal here was to model the shape and variability of change in SEI scores over time among middle and high school students. Variables included in these models are described in Table 2.3.

Table 2.3
Operational definitions and descriptive statistics for variables studied

	Variable	Description
Grouping variable	id (j)	contrived unique student identifier
Dependent variable	SEI (y_{ij})	SEI total score, equal to the mean score across all five SEI factor scores
Independent variables	time (t_{ij})	time in years from the start of Grade 6
	male (w_j)	dummy-coded variable for gender (1: male, 0: female)

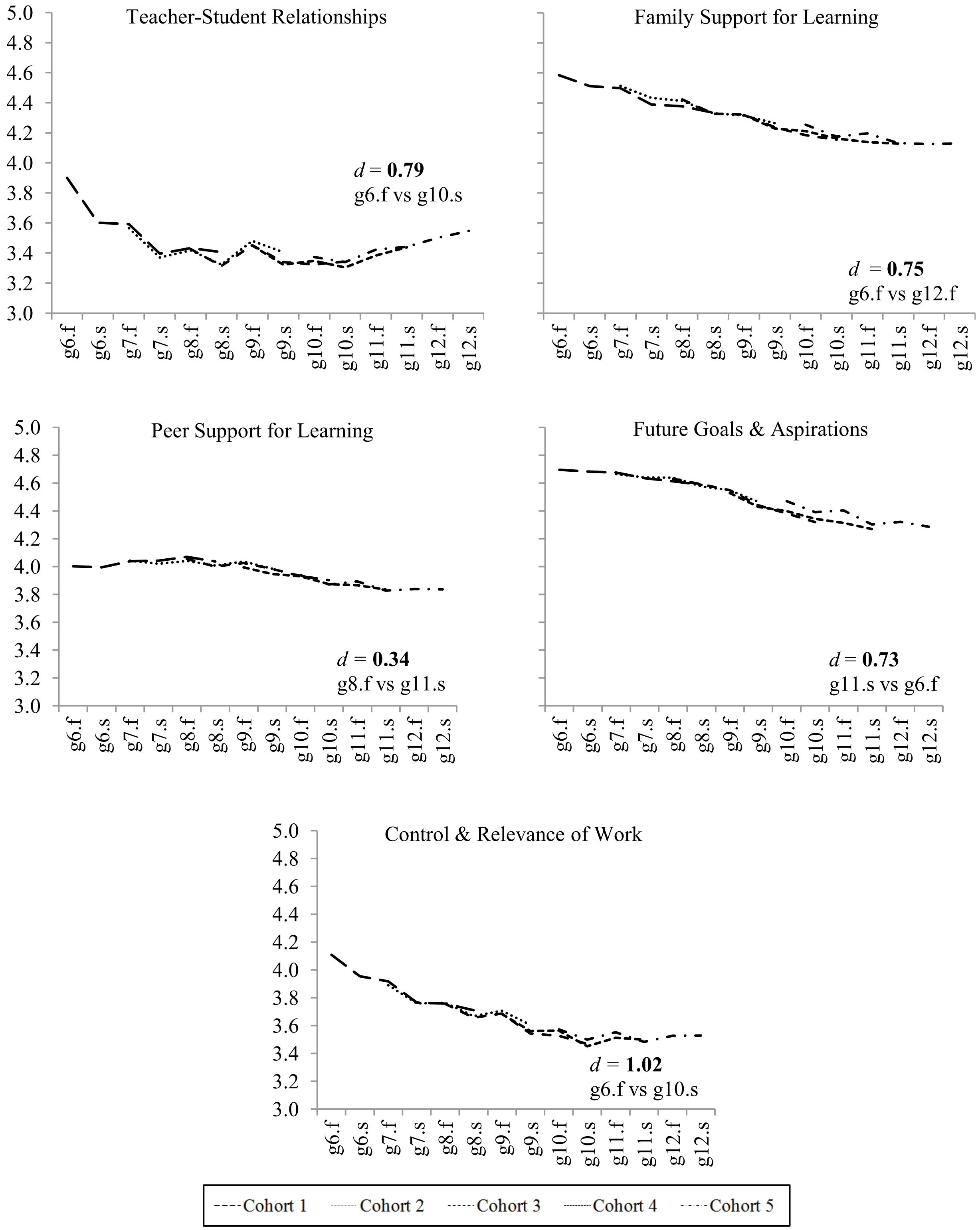

Figure 2.4 Mean-level change of SEI factors by grade and cohort. Semester-to-semester change in means in grades 6–12, using an accelerated longitudinal design. The x-axis represents the grade level and semester for the SEI score. The y-axis represents cohort SEI means and was purposely rescaled from a 1–5 to a 3–5 scale to provide a closer view of how well cohort trajectories overlapped.
g = grade f = fall s = spring *d* = Cohen's *d* statistic

Model 1: Basic quadratic model. Plots of SEI score means through middle and high school appeared curvilinear in most cases, so modeling started with the inclusion of a quadratic term for time (time^2). It was expected that students would vary considerably in terms of their initial engagement level as well as in their rate of growth, and exploratory analysis showed that males and females differed in average engagement level at any given grade level. Model 1 was built with these ideas in mind:

$$y_{ij} = \beta_1 + \beta_2 w_j + \beta_3 t_{ij} + \beta_4 {t^2}_{ij} + \zeta_{1j} + \zeta_{2j} t_{ij} + \epsilon_{ij}$$

where y_{ij} was the SEI score, t_{ij} was the year since Grade 6 of student *j* at occasion *i, w_j* was a dummy variable for gender, and ζ_{1j} and ζ_{2j} were a random intercept and random slope, respectively. The occasion specific residual ϵ_{ij} allowed deviation in SEI scores y_{ij} from the quadratic trajectories. Model 1 estimates are shown in Table 2.4.

All coefficients for Model 1 were significant at the $p < .05$ level. Estimates indicated that males' self-reported engagement was roughly a tenth of a point lower on the SEI on average than that of girls in the same grade. The growth curve trended downward at time 0 (negative coefficient) but the rate of decline lessened over time (just slightly positive coefficient for the squared term). This trend is illustrated in Figure 2.5, which also shows the difference in intercepts between genders. The estimated random-intercept standard deviation of 0.36 points suggested considerable variation between students. The average decrease, however, in SEI score per year varied with a standard deviation of just under a tenth of a point per year. In combination, these findings suggest that the vast majority of students gradually decreased over time but that their starting points in the trajectory varied to a larger degree.

Table 2.4
Quadratic growth model estimates of SEI total score change from grade 6 to 12

	Model 1: Basic Quadratic	Model 2: Interaction	Model 3: Complex L1 Variation		Model 4: Complex L1 & L2 Variation	
	β (*SE*)	β (*SE*)	β (*SE*)		β (*SE*)	
Fixed effects						
intercept	4.40 (.007)	4.41 (.008)	4.41 (.007)		4.41 (.007)	
male	-0.09 (.004)	-0.10 (.008)	-0.09 (.004)		-0.09 (.004)	
male×time	--	0.00 (.002) †	--		--	
time	-0.11 (.003)	-0.11 (.004)	-.11 (.003)		-0.11 (.003)	
$time^2$	0.004 (.000)	0.004 (.000)	0.004 (.000)		0.005 (.000)	
Random effects					*F*	*M*
$\sqrt{\psi_{11}}$	0.36 (.007)	0.36 (.007)	0.36 (.006)		0.36 (.008)	0.35 (.011)
$\sqrt{\psi_{22}}$ [time]	0.08 (.002)	0.08 (.002)	0.08 (.002)		0.08 (.003)	0.09 (.003)
ρ_{21}	-0.43 (.026)	-0.42 (.026)	-0.43 (.025)		-0.47 (.030)	-0.38 (0.042)
			F	*M*		
$\sqrt{\theta}$	0.31 (.001)	0.31 (.001)	0.28 (.001)	0.34 (.002)	.28 (.001)	.34 (.002)
Log likelihood	-64469.72	-64468.29	-63932.71		-63846.76	

†All estimates significant at the $p < .001$ level, except for cross-level interaction in Model 2 between time and gender ($p = .091$), which was not retained in subsequent models. Model 3, shaded in grey, was considered the final model.

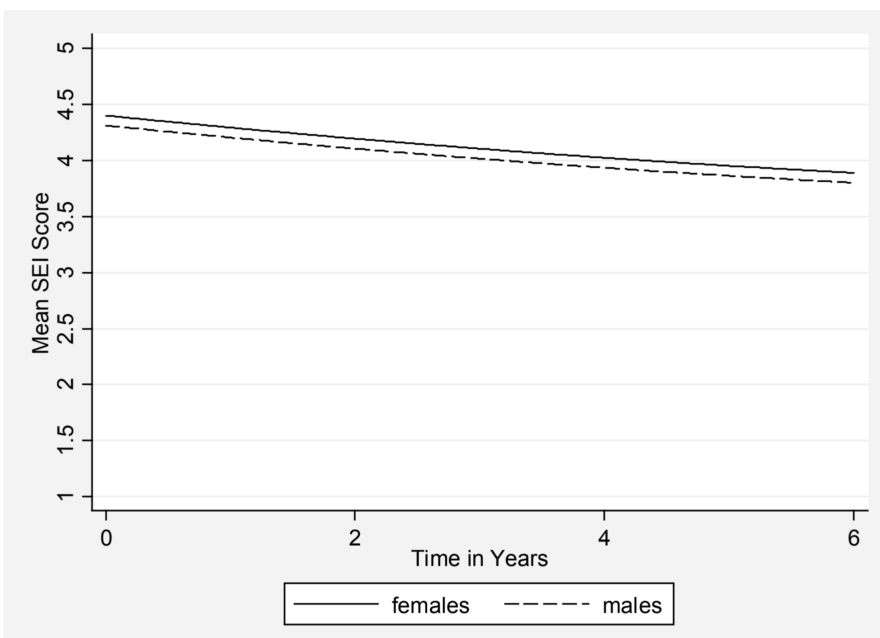

Figure 2.5: Model 1 based normative trajectory of SEI from Grade 6 to 12.

Next, Model 1 was reformulated using a two-stage approach, as described in Raudenbush and Bryk (2002) and Rabe-Hesketh and Skrondal (2012), to create the remainder of random-coefficient models in Models 2 through 4. The level-one model was specified with student-specific intercept π_{0j} and slope π_{1j} coefficients:

$$y_{ij} = \pi_{0j} + \pi_{1j}t_{ij} + \pi_2 t^2{}_{ij} + \epsilon_{ij}$$

And the level-two model was then formulated with the intercepts and slopes as outcomes:

$$\pi_{0j} = \gamma_{00} + \gamma_{01}w_j + r_{0j}$$

$$\pi_{1j} = \gamma_{10} + r_{1j}$$

where gender (w_j) is a covariate only in the intercept equation, r_{0j} and r_{1j} represent the residuals, and where π_{0j} and π_{1j} represent the random effects.

Model 2: Quadratic model with a cross-level interaction. In Model 2, a cross-level interaction ($\gamma_{11}w_j t_{ij}$) was added between gender and time by including the term **male** in the level-two model for π_{1j}, adding $\gamma_{11}w_j$. The estimate for this interaction was not significant at the $p < .05$ level, so this term was dropped in Models 3 and 4 (see Table 2.4).

Model 3: Quadratic model with random intercept by gender. Models 1 and 2 assumed that the random intercept, random slope, and level-one residual were all homoskedastic—i.e., variance was constant for all students for these effects. Models 3 and 4 allowed complex variation for the level-1 residuals and for random effects, exploring whether variances should be allowed to depend on gender. First, in Model 3, the level-one residual variance θ was freed to differ between males and females by adding gender-specific paramaters $\theta^{(F)}$ and $\theta^{(M)}$. All coefficients previously in the model remained significant at the $p < .05$ level, and intercept and slopes changed only slightly from Model 1. The point estimates for the standard deviation of the level-one residuals were .34 and .28 for males and females, respectively, with 95% CIs that did not overlap. In other words, it appeared that males were more likely to deviate from the normative trend than females. This finding paired with the considerable likelihood ratio improvement from Model 1 to 3 indicated that allowing this residual variance to differ by gender resulted in better model fit.

Model 4: Quadratic model with random intercept and slope by gender. In Model 4, level-one residual variance was continued to be allowed to differ by gender and, additionally, the random-intercept variance ψ_{11} and random slope variance ψ_{22} and their covariance ρ_{21} were freed to differ by gender, adding the parameters ${\psi_{11}}^{(F)}$, ${\psi_{22}}^{(F)}$, and ${\rho_{21}}^{(F)}$ for females and ${\psi_{11}}^{(M)}$, ${\psi_{22}}^{(M)}$, and ${\rho_{21}}^{(M)}$ for males. While freeing these parameters resulted in differences in covariance between genders, random intercept and slope variances were nearly identical. Further, the likelihood ratio improvement from Model 3 to 4 was not nearly as substantial as the shift from Model 1 to 3. Therefore, Model 3 was retained as the final model.

Individual change

The RCI was used to quantify the probability of observing a student's difference score over a two-year increment equal to or greater than the one observed, assuming that no change had occurred. The RCI adjusts for the estimated unreliability of measurement, providing a useful method of separating true change in SEI factor scores from change due to measurement error (Robins et al., 2001). Calculations of the RCI, which required a retest correlation estimate, were informed by the observed annual retest coefficients found in the stability analyses above. Essentially, S_{diff} scores represented the distribution of change scores if change were due solely to measurement error. In this way, assuming normality, RCI scores greater or less than 1.96 should have occurred 5% of the time if change were to happen by chance alone (2.5% less than -1.96, 2.5% greater than +1.96; Blonigen et al., 2006). Because the assumption of normality was particularly important here, RCIs were computed for SEI total scores rather than for subscale factor scores. Results of comparisons between Grades 6 and 8, between 8 and 10, and between 10 and 12 are presented in Table 2.5.

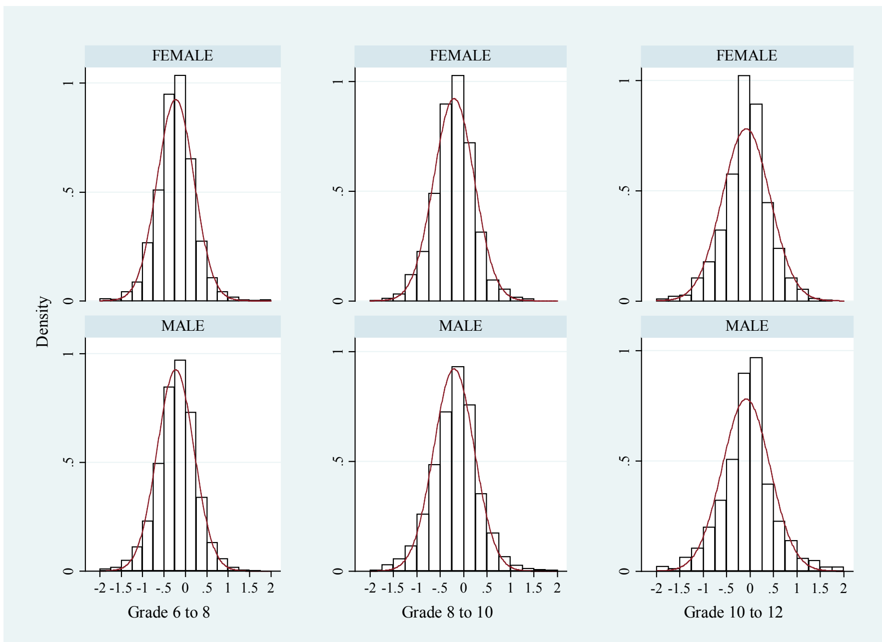

Figure 2.6: Distributions of SEI raw score differences. Comparisons over two year intervals, with normal density curves plotted in red.

Chi-square tests were used to examine the distribution of students observed to have reliably changed compared with the distribution of students expected to have RCI scores ±1.96 due to chance alone. One two-year comparison for males, Grades 10 to 12, was not significant at the $p < .05$ level. All other comparisons were highly significant for males and females, an indication of overall reliable change in SEI scores. In most cases, students showing reliable decreases in cognitive/affective engagement occurred 2 to 3 times more often than would be expected by chance alone. Together with the retest stability and mean-level findings, these results suggest that—while overall cognitive/engagement appears moderately stable and to decrease gradually for most students over time—there is also the possibility of distinct trajectories suggested by the data, particularly for decreasing trajectories that are much more rapid than for most students.

Table 2.5
SEI reliable change index distributions by gender and grade

Grade Comparison	*n*	Decreased (%)	Same (%)	Increased (%)	χ^2
Grades 6 to 8					
Male	2,872	6.5	92.4	1.1	65.34***
Female	2,851	7.3	91.6	1.2	82.12***
Total	5,723	6.9	92.0	1.1	149.34***
Grades 8 to 10					
Male	2,196	6.0	92.7	1.3	39.99***
Female	2,310	7.1	91.5	1.4	60.02***
Total	4,506	6.5	92.1	1.4	99.13***
Grades 10 to 12					
Male	1,281	3.9	93.3	2.8	4.37
Female	1,389	4.9	93.0	2.1	11.67*
Total	2,670	4.4	93.1	2.4	14.79**

Note. Decreased (%), Same (%), and Increased (%) refer to the percentage of individuals whose SEI total score decreased, remained the same, or increased, respectively, according to the RCI. Chi-square tests ($df = 2$) compared the observed distribution of changers and nonchangers to the expected distribution if changes were due to chance alone.
* $p < .01$ ** $p < .001$ *** $p < .0001$

Discussion

In this study, an accelerated longitudinal design was used to investigate the developmental trends of cognitive and affective engagement as measured by the SEI (Appleton et al., 2006). Retest correlations showed that all factor scores on the SEI were moderately stable year to year through middle and high school, with CRW showing more stability than expected and estimates for other factors falling very close to hypothesized stability levels based on Conley's (1984) stability hierarchy. Mean-level analyses revealed relative continuity in PSL over time, while the other four factor scores tended to show gradual decline for the majority from grades 6 to 12. These analyses also demonstrated a commonly occurring semester-to-semester rise and fall in mean estimates, likely a manifestation of the strong influence of contextual change on engagement. Other indications of contextual influence were the initial significant decline in TSR scores at the start of middle school and the apparent accelerated decline in scores during the first year of high-school on most factors. On the former point, it is worth noting, however, that the 6th grade estimated mean for TSR was based on one estimate, and thus confidence in the accuracy of that estimate is lower than for other grade-to-grade comparisons. The 9th grade accelerated decline, on the other hand, was visible in many factors and across cohorts, suggesting that the transition between middle and high school constitutes a change in context associated with slightly lower levels of cognitive/affective student engagement for most students. In individual-level analyses, data indicated that at least a small portion of the student population followed change trajectories that differed substantially from the normative group—highly reliable evidence of change because measurement error was controlled for in these analyses.

Developmental studies of academic achievement motivation, as noted in Skinner and Pitzer (2012), provide a broad empirical context for interpreting the results of the present study. Similar to the results of this study, academic achievement motivation and engagement research suggest that mean-levels of engagement decline across the school years (Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006) and that an individual's engagement is correlated over time: from the beginning to the end of the school year (Skinner & Belmont, 1993), from one grade to the next (Gottfried, Fleming, & Gottfried, 2001), and between middle and high school grades (Gottfried et al., 2001; Marks, 2000; Otis, Grouzet, & Pelletier, 2005). The present study indicated that the longitudinal aspects of the SEI are consistent with this broader framework.

Longitudinal studies in which student engagement is explicitly studied as a multidimensional construct over a long period of time, however, are relatively scarce in the literature. One longitudinal study, conducted by Wylie and Hodgen (2012), and another accelerated longitudinal study, conducted by Janosz et al. (2008), provide further valuable context due to their similarity to the present study. Although these studies did not investigate the stability of engagement, they did explore its group-level and individual-level change, the results of which are also commensurate with findings here for the SEI. In the study by Janosz et al., growth mixture modeling revealed several distinct normative and several non-normative trajectories of engagement, which was measured with a composite variable consisting of cognitive, affective, and behavioral indicators. The present results of the growth model analysis of the SEI—which found considerable variation in intercepts between students overall yet much less variation in slopes—and the RCI analysis—which found that less than 10% of the sample showed reliable evidence of considerable change over a two-year period—fit well with the work of Janosz et al., who found that 91% of their sample followed one of the normative trajectories that showed gradual decline over time. These normative trajectories were differentiated by intercepts of high, moderate, and low engagement. The other 9% in their study followed unstable trajectories consisting of either variable change over time or rapid decline. The growth curve and RCI analysis in the present study were valuable first steps in checking for reliable evidence of group- and individual-level change in SEI scores beyond what might occur simply by chance. Future studies should further investigate latent normative and non-normative trajectories of the SEI, or of SEI factors in combination with behavioral engagement data.

There are several limitations of this study that warrant consideration. The first is conceptual, in that the theoretical perspective of this study represents one of many varying perspectives of student engagement, of its definition and composition, and of how it should be measured. A prominent methodological limitation was missing data, largely due to attrition from students leaving the district. The accelerated design was incorporated to limit the effects of attrition, but this was only useful to the degree that each cohort consisted of a reasonably representative continuation of the previous cohort. One possible reason for the plateau or gradual increase in some of the factor scores in the later grades is that, once the legal dropout age was reached, the later cohorts were less likely to be representative of the true population. In other words, if the students who had left the district had been there to fill out the SEI in grades 11 and 12, the means for these grades would have perhaps been lower.

The results of the present study have practical implications, particularly for applied researchers using or intending to use the SEI as a measure of engagement. The first is that the consistency between the results of this study and the larger framework of research on engagement can now be added to the growing body of evidence supportive of the SEI's construct validity. The other major implication for practice is that change in a student's factor score—

whether for the total score or for a subscale—should be interpreted with the annual stability of the measure in mind. And due to the moderate levels of this stability, in terms of interpreting scores, the best practice may be to consider developmental trends in a student’s scores over longer periods of time, such as from early to late middle school, rather than examining scores in isolation or from one semester to the next. If future studies of the SEI find sound evidence of the variety of engagement trajectories described by Janosz et al. (2008) and Wylie and Hodgen (2012), a valuable next step would be to understand what these trajectories—particularly early trajectories—can tell us about who is falling off the path to graduation and college and career readiness.

CHAPTER 3

INCREMENTAL VALIDITY OF COGNITIVE AND AFFECTIVE ENGAGEMENT OVER SCHOOL RECORD DATA[2]

[2]Lovelace, M.D., A.L. Reschly, and J.J. Appleton. To be submitted to *Journal of Psychoeducational Assessment*.

Introduction

Research on student engagement with school has been active for over 25 years and is conducted by scholars from a variety of disciplines and nationalities (Christenson, Reschly, & Wylie, 2012). Interest in the construct extends well beyond academia, as student engagement has also become a growing area of focus for many educators, practitioners, and policymakers (Appleton, Christenson, & Furlong, 2008; Fredricks, Blumenfeld, & Paris, 2004; National Research Council and the Institute of Medicine, 2004). There are many reasons for this. Student engagement is useful for understanding dropout and for promoting school completion (Christenson et al., 2008; Finn, 2006). It is associated with academic, social, and emotional learning outcomes (Fredricks et al., 2004); and along with academic achievement, indicators of engagement are some of the strongest predictors of high school dropout (Janosz, 2012; Rumberger & Lim, 2008). Further, models of engagement are typically multidimensional and rooted in context (Appleton et al., 2008; Christenson et al., 2012), which allows for richer characterizations of students as cognitively, emotionally, and behaviorally complex individuals, who are developing in, influencing, and being influenced by, a complex ecology. Additionally, engagement is a practical centerpiece for intervention efforts because, as an alterable variable (Lehr, Johnson, Bremer, Cosio, & Thompson, 2004), it is capable of being shaped by context (Reschly & Christenson, 2006). Further, the number of evidence-based interventions and strategies for encouraging student engagement is growing (Christenson et al., 2012). Finally, there are a variety of engagement measures with adequate psychometric properties available for use in both research and school settings (Fredricks et al., 2011).

This last reason, increased availability and awareness of engagement measures, is a point of focus in the present paper. A significant motive for developing engagement instruments is to

.

use them in schools for data-based decision-making (Appleton, 2012). This is because monitoring engagement may provide stakeholders with the opportunity to respond proactively to students or school issues most in need of intervention, by helping them to target school-wide issues or to identify at-risk students (Appleton, Reschly, & Martin, 2013). But the ability to monitor student engagement and to evaluate the effectiveness of efforts to improve it requires reliable and valid measures (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

A variety of instruments have been developed to measure engagement, many of which have demonstrated adequate psychometrics in a variety of areas of reliability and validity (see Fredricks et al., 2011 for a review of 21 measures); nevertheless, questions remain about some aspects of these instruments' properties, particularly with regard to their use in longitudinal studies and to their practical value for school use. A primary impetus for this paper is the notion of using multidimensional engagement measures for data-based decision-making in a school setting, which makes one of these measures, the Student Engagement Instrument (SEI; Appleton, Christenson, Kim, & Reschly, 2006), suitable for the focus of the present study. In comparison with other available engagement measures, the SEI is one of the few student engagement instruments that is comprehensive (includes a variety of subscales related to cognitive and affective engagement; Fredricks et al., 2011), based on a theoretical model, available for free (published in full in Appleton et al. 2006 and in Betts, Appleton, Reschly, Christenson, & Huebner, 2010), and has shown wide-spread applicability, being used in districts across the U.S. and internationally (e.g., Moreira, Vaz, Dias, & Petracchi, 2009). The SEI has been used both as a research tool in several studies of engagement (e.g., Lewis, Huebner, Reschly, & Valois, 2009; Reschly, Huebner, Appleton, & Antaramian, 2008) and in practice (Appleton, 2012; Appleton,

.

Reschly, & Martin, 2013). Gwinnett County Public Schools in Georgia has used the SEI continuously since 2007 to provide educators in the district with information that may supplement existing data on student progress in school (Appleton, 2012; Fredricks et al., 2011).

Use of the SEI as a tool for data-based decision-making raises validity questions that are yet to be answered. That is, although findings from previous validation studies on the SEI (Appleton et al., 2006; Betts et al., 2010; Lovelace, 2010, 2011) support a variety of aspects of its reliability and validity, one type of validity highly relevant to data-based decision-making in need of further study for the SEI is *incremental validity*. This is a practical aspect of validity referring to whether a measure adds to the prediction of a criterion or outcome beyond what can be predicted by other sources of data (Sechrest, 1963; Hunsley & Meyer, 2003). Although the concept of incremental validity is relatively straightforward, assessment of it in applied settings can be complicated. In clinical psychology, for instance, improvement in prediction can mean different things, such as increased power, sensitivity, specificity, or—for decision-making judgments—efficiency of prediction beyond what is generated on the basis of other data (Hunsley & Meyer, 2003). This latter definition, efficiency of prediction—a designation in line with the earliest meaning of incremental validity—is what the present study seeks to address with regard to the SEI. Sechrest (1963), who first articulated the concept, argued that for a measure to have true utility in applied settings, it should demonstrate incremental validity over readily available data (e.g., interview or case history data). Although Sechrest was writing with an audience of clinical psychologists in mind, his perspective on incremental validity also applies to school settings. School districts have a number of clearly defined prediction tasks (such as academic risk, dropout, college and career readiness), an abundance and variety of data at their disposal, and considerable incentives for optimizing the prediction of outcomes (reducing costs,

.

using resources more effectively, meeting accountability standards). In practical terms, envisioning Sechrest's argument from the perspective of a school psychologist, principal, or superintendent considering whether to use the SEI in his or her school, the question of incremental validity could become: *Do student scores on the SEI tell me anything more about student and school outcomes than what I could have gathered from information already available in school records?* In this sense, incremental validity is a relatively strict test of the SEI because it demands not only that the measure predict an outcome better than what could occur by chance alone, but that it also shows additional explanatory value relative to less expensive sources of information. The broad objective of this study was to test the incremental validity of the SEI in terms of the following research question: Does the SEI incrementally predict on-time graduation or dropout when controlling for relevant data commonly available in school records?

Method

Participants and Procedures

Data were provided by the Office of Research and Evaluation in Gwinnett County Public Schools (GCPS), a school district located in the metro Atlanta area. GCPS is particularly suited to studying the incremental validity of the SEI because it has administered the measure since September 2007 as part of a district-wide student advisement program (Appleton, 2012). A cohort was drawn from GCPS data, consisting of first-time ninth graders who had been enrolled in the district the previous year (N = 10,067). Students were included if they were enrolled for ≥ 65% of the academic year (i.e., ≥ 117 days) in 2007-2008. These students represented 15 high schools, with 2007-2008 school enrollments ranging from 425 to 1,002 students. With four academic years of SEI data available, studying this cohort longitudinally allowed for analyses of the relationships of various student and school characteristics with dropout and graduation.

.

Table 3.1
Student and school demographics (K-12): District to state comparison

	GCPS (%)	Georgia (%)
Ethnicity		
White	36	46
Black	27	38
Hispanic	22	10
Asian	10	3
Other	5	3
Students with Disabilities	11	11
Limited English Proficient	15	5
Eligible for Free/Reduced Meals	41	51

Source: Office of Student Achievement Report Card, www.gaosa.org

Measures and Covariates

Dropout and on-time graduation. De-identified data supplied by GCPS were used to construct a series of student-level variables for the present study. On-time graduation and dropout were the primary outcomes of interest. *Dropout* was defined as leaving high school before the end of the observation (September 2011) for any reason besides graduation, transferring out of the district, or earning a certificate of completion. Students who died or left due to a serious illness or accident were not included in the data. *On-time graduation* was defined as graduating with a full high school diploma by the end of summer 2011. Students who earned a certificate of completion or a special education degree were identified as *completers* but not as graduates.

Cognitive/affective engagement. Cognitive and affective engagement variables were constructed for each semester from student responses to items on the SEI. The SEI is a 33-item survey that was designed to measure a student's self-perception of their cognitive and affective engagement in school (Appleton et al., 2006). Administration of the SEI included standardized

.

procedures, including items being read aloud to students to limit unwanted effects from variation in reading levels (Appleton, 2012). The SEI measures five facets of cognitive and affective student engagement, namely: Teacher-Student Relationships (TSR), Control and Relevance of School Work (CRW), Peer Support for Learning (PSL), Future Aspirations and Goals (FGA), and Family Support for Learning (FSL). TSR, PSL, and FSL were formulated to measure aspects of affective engagement, whereas CRW and FGA are related to cognitive engagement. In the 2007-2008 school year, the SEI used a four-point Likert response format ranging from 1=*Strongly Disagree* to 4=*Strongly Agree*. From 2008-2009 onward, the scale was changed to a five-point format with the addition of a *Neither Agree Nor Disagree* option anchored at 3. The SEI was coded so that higher scores signified higher levels of engagement. Using the same logic as Betts et al. did in their 2010 validation study of the SEI, two SEI items comprising a sixth factor, Extrinsic Motivation (EM), were not included in these analyses. A 5- rather than a 6-factor model was included in the present study because the two items that load onto EM are both negatively worded and the factor is underdetermined (Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005).

Empirical evidence from validation studies of the SEI support its use as a cognitive and affective student engagement measure. Good internal consistency estimates (α = .72 to .88 across factors) have been reported in two studies (Appleton et al., 2006; Betts et al., 2010). Exploratory and confirmatory factor analysis methods have suggested that the SEI has a meaningful factor structure (Appleton et al., 2006; Betts et al., 2010), and structural equation modeling techniques have revealed that its factor structure is stable across cross-sections of grades six through twelve (Betts et al., 2010). Further, as the findings above in Study 1 suggest, the longitudinal

.

performance of the SEI fits well with theory and with what would be expected given prior empirical evidence from other longitudinal studies of engagement and achievement motivation.

Student data common in school records. The remaining variables were constructed from data commonly kept in school records, which were grouped into several broad categories: demographic characteristics, academic achievement, and behavioral disengagement. These variables were chosen based on what data were available in the district data stores and according to comprehensive empirical evidence on associations between student variables and school completion found in a 2008 literature review by Rumberger and Lim, which covered 25 years of dropout research. Over-age, an indicator related to retention, is typically defined as being 1 or 2 years older than classmates (Rumberger & Lim, 2008). Here, ages were known only within a 3-month interval; to account for this uncertainty, over-age was defined as being at least 1.5 years older than the average age of classmates. Several status variables, like ethnicity and special education status, are known to associate with differences in school outcomes as individual indicators, but these effects have commonly been shown to be insignificant once variables related to achievement, socio-economic status, and behaviors are taken into account (Rumberger and Lim, 2008). These demographic characteristics were still included in the predictive modeling process, however, as a check on how well the results of this study match with the literature. Gender, which many studies have found to be unrelated to school completion once other variables are controlled for, was included because, overall, the evidence on its associations is mixed (Rumberger & Lim, 2008). Operational definitions and descriptive statistics for all variables are provided in Table 3.1.

.

Analysis Procedures

Predictive efficiency of grade nine indicators. The analytic logic for this study was inspired by the work of Gleason and Dynarski (2002) and Balfanz, Herzog, and Mac Iver (2007), who showed that, to be useful for predicting dropout or on-time graduation, variables of interest need to demonstrate high predictive power and yield. Balfanz et al. assessed the utility of dropout predictors with a two-pronged approach, requiring (1) high positive predictive value and (2) high sensitivity. Positive predictive value (*PPV*)—referred to as predictive *power* by Balfanz et al—refers to the proportion of students for whom the event was true among those identified by the predictor (i.e., $PPV = \frac{true\ positives}{true\ positives + false\ positives}$). Sensitivity—referred to as predictive *yield* by Balfanz et al.—is indicated by the proportion of the target group (e.g., all dropouts) that the predictor correctly identified (i.e., $Sensitivity = \frac{true\ positives}{true\ positives + false\ negatives}$). To demonstrate high predictive efficiency, each of the student-level variables involved in this study was subjected to a dual-criterion test:

(1) PPV rate is at least double the cohort dropout rate (or, in the case of on-time graduation, at least half the rate), and

(2) Identifies a substantial enough portion of the true target group to be valuable in intervention efforts. A sensitivity estimate of at least .05 (i.e., 5% of the total target group) was determined to be a reasonable minimum criterion for a single predictor.

.

Table 3.2
Operational definitions and descriptive statistics for variables studied

	Variable	Description	*M*	*SD*	*SE*
Dependent variables (y_{ij})	Dropout	1=Identified as a dropout; 0=Not identified as a dropout.	.05	.21	.002
	Graduate	1=Graduated with a full diploma from the district within 4 academic years; 0=did not.	.67	.47	.004
Grouping variables	Contrived_id (*i*)	Unique contrived student identifier			
	School_id (*j*)	Contrived identifier of enrolled high school in 2007-2008			
Independent variables	Female	1=Yes; 0=Male	.50	.50	.005
Demographic data	*Race/Eth:* Black	1=Yes; 0=No. White was the reference category in regression analyses.	.25	.43	.004
	Hispanic	1=Yes; 0=No.	.21	.40	.004
	Other	1=Asian/Pacific Islander, Native American/Alaskan, or Multiracial; 0=None of these.	.15	.36	.004
	White	1=Yes; 0=No.	.39	.49	.005
	FRL	1=Identified as eligible for free or reduced-price lunch in 9th grade; 0=Not identified.	.37	.48	.005
	Over-age	1=At least 1.5 years older than typical age at start of 9th grade; 0=Not over-age	.02	.15	.002
	Special Ed.	1=Identified as receiving special ed. services as of start of 9th grade; 0=Not identified.	.13	.33	.003
	Spanish	1=Primary language is Spanish; 0=Primary language is another language	.16	.37	.004
	Gifted	1=Identified in records as Gifted; 0=Not identified	.19	.39	.004
Academic achievement	Achievement[a]	Prior achievement; mean of z-scores (based on state *M* and *SD*) across Grade 8 CRCTs for Math, Reading, and English/Language Arts.	.37 (.34)	.90	.009 (.009)
Behavioral disengagement	OSS	1=At least 1 out-of-school suspension (OSS) in 9th grade; 0=No OSS	.14	.34	.003
	Attendance	Percentage of enrolled days attended in 9th grade.	96.2	5.10	.051
Cognitive/affective engagement	TSR[a]	Average factor score for TSR on the SEI for grade 9.	2.76 (2.75)	0.48	.005 (.005)
	PSL[a]	Average factor score for PSL on the SEI for grade 9.	3.20 (3.19)	0.45	.005 (.004)
	FSL[a]	Average factor score for FSL on the SEI for grade 9.	3.43 (3.42)	0.48	.005 (.005)
	FGA[a]	Average factor score for FGA on the SEI for grade 9.	3.61 (3.60)	0.43	.004 (.004)
	CRW[a]	Average factor score for FSL on the SEI for grade 9.	2.93 (2.91)	0.45	.005 (.004)

Note. The superscript [a] denotes that a variable had missing data and that analyses with these variables involved multiple imputation. Mean (*M*), standard deviation (*SD*), and standard error (*SE*) shown for non-imputed data. For any variable involving imputation, *M* and *SE* following multiple imputation are shown in parentheses below the estimates.

Multilevel logistic regression. The dual-criterion test was intended to explore and assess the predictive utility of individual 9th grade predictors. It was possible, however, that the explanatory effect of some of these predictors would overlap. If FGA were to demonstrate predictive utility for identifying non-graduates, for example, it is possible that every non-graduate identified by FGA would have already been correctly identified by another variable that was easier to gather, like attendance, because it was already in school data. Further analyzing these predictors within the framework of multiple regression allowed for a statistical analysis of the incremental validity of an SEI factor score, or its unique explanatory value, over a wide variety of school data. Estimates of student-level and school-level effects on dropout and graduation were analyzed through multilevel logistical regression modeling using Stata software (StataCorp, 2011). The data had an inherent two-level structure of students i (n=10,067) nested within high schools j (n=15). Model building for the separate analyses for the outcomes of dropout and on-time graduation were carried out in a parallel fashion, with the same essential model building sequence in both sets of analyses, including the same predictors.

In a series of separate two-level, random-intercept logistic regression models, the response variables for on-time graduation and dropout were each regressed on a variety of student-level variables. For each response variable, a series of four primary models were built, representing (1) an *unconditional model* with no level-one or level-two predictors, (2) a *demographic data model*, with a variety of level-one status variables common in school data (e.g., free/reduced lunch eligibility, ethnicity), (3) a *demographic and academic data model*, with a level-one prior achievement variable added, (4) and a *demographic, academic, and behavior data model*, with level-one indicators for attendance and out-of-school suspensions added. Originally, analyses were to include level-two predictors (e.g., percentage of students eligible for

.

free-/reduced-price lunch), but school-level variance was found to be insignificant once powerful level-one variables were added, thus incremental analyses did not progress past the addition of level-one variables. In each step, an additional subset of models were run, first adding an SEI factor, then removing it and adding another SEI factor, eventually running the model with all SEI factors included. This progression allowed for the assessment of the incremental contribution of new variables, including SEI variables at each step. Note that, as further explained in the results below, data exploration, as well as the PPV and sensitivity analyses, were conducted prior to the inferential analyses to better inform which factors to include in the incremental validity analyses.

Statistically significant incremental validity was assessed by examining logistic regression coefficients (exponentiated as odds ratios) and their 95% CIs for each SEI factor included in the model. For effect size of overall model prediction, a pseudo-R^2 statistic was calculated, as recommended by Peugh (2010), by taking the square of the correlation between the predicted conditional probabilities and the outcome variable. The difference in R^2 after new variables were added was used as an absolute index of effect size of the validity increment. Rather than interpreting this effect in traditional terms, such as by Cohen's benchmarks (Cohen, 1992), guidelines by Hunsley and Meyer (2003) were used instead that were proposed specifically for tests of incremental validity. Based on Nunnally and Bernstein's (1994) observation that increases in R^2 are generally small in social science research by the time a third substantial predictor has been entered into a regression, Hunsley and Meyer suggested that a lower R^2 difference, such as .0225 (square root = .15), would constitute a reasonable contribution to the regression. Although this is a small effect by traditional standards, these criteria were considered to be stringent in the present study given the many substantial controls included at

.

each step, and that, even within the second model, SEI factors would be added well after a third meaningful variable was included in the model.

Because dropout was confined to a strict definition of only those students identified as dropouts by the district, it was relatively rare in the studied population (< 5%). When rare events are binary they can be difficult to analyze for reasons both practical (e.g., costs of gathering extensive data on thousands of participants to capture rare event data) and statistical (King & Zeng, 2001; Lacy, 1997). Issues related to efficient data-gathering obviously did not apply here, but statistical issues were a concern because logistic regression often underestimates event probabilities when nonevents greatly outnumber events (King & Zeng, 2001). Unbiased estimates can be achieved, however, through careful design and post-estimation statistical correction (King & Zeng, 2001). For the dropout models, where 1s were greatly outnumbered by 0s, a *case-cohort* design was used, in which a balanced random sample was constructed by selecting students on the dependent variable, and then, following estimation, prior correction (King & Zeng, 2001) was applied to the intercept to account for the case-cohort sampling design. SEI scores and other variables were available beyond 9th grade, but only 9th grade variables were used in sensitivity, PPV, and regression analyses to maintain a practical perspective.

.

Results

Descriptives and Missing Data Analysis

Four years later. Figure 3.1 displays outcomes for the ninth grade cohort in terms of graduation status proportions as of the beginning of September 2011, one semester after the expected graduation date.

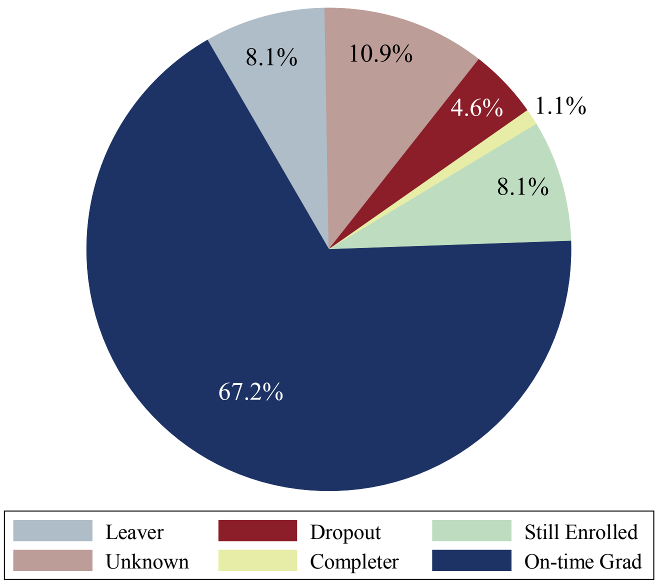

Figure 3.1: Four-year outcomes.

These descriptive statistics indicate that about one fifth of the cohort's graduation status at the end of the study was unknown, either because the student transferred and never re-enrolled or because there was no record of their graduation status and the student did not re-enroll as of September 2011. Without knowing the outcomes for these students, somewhere between 13.8% and 32.8% of the cohort did not earn a high school diploma within four years. A cross-tabulation of leave year by leave reason indicated that transferring out of the district for known or unknown reasons did not occur at a consistent rate from grade 6 to 12, but the trends for these categories did not also mimic the clear linear pattern that dropout proportions seemed to follow.

.

Data on demographics, behavior, and attendance were complete for all cases in the cohort for grade nine, but data for prior achievement and SEI responses were not. The difference between SEI missingness and the overall leave rate in 9th grade—what could be considered the expected proportion of missingness—indicated that roughly 10% of SEI data was missing beyond what naturally could be expected due to attrition in the first year. When there is no systematic pattern of missingness, both observed and unobserved, data are considered missing completely at random (MCAR; Schlomer, Bauman, & Card, 2010). A less stringent assumption of missingness is when data are considered missing at random (MAR), referring to systematic missingness that can be accounted for by other data in the dataset. Engagement theoretically should influence a student's propensity to be present to complete a survey of any kind, let alone an engagement survey, and this missingness should be accounted for by covariates in the data. MAR is an assumption that is essentially not able to be proven, but MAR was considered a reasonable assumption for the missingness in the SEI and achievement data because the breadth and completeness of surrounding data allowed for a multiple imputation model that included a wide variety of covariates relevant to potential underlying reasons for missingness (such as absence rates, disciplinary records, prior achievement, grade level, prior and later engagement scores, outcome status). Due to this assumption and the low rate of missingness overall, multiple imputation methods (Little & Rubin, 2002) were used to generate a universe of 5 complete datasets. These imputations were generated with the multivariate imputations by chained equations (MICE) procedure using Stata software (StataCorp, 2011).

Exploratory Data Analysis

Graphical exploration of the data with respect to associations of SEI factors to four-year outcomes suggested consistent directional trends (i.e., higher reported engagement associated

.

with higher graduation rates and lower dropout rates) but also varying levels of strength across factors (see Figure 3.2). For instance, as Figure 3.2 shows, the prevalence of dropout in the data among students with lower scores on FGA in 9th grade was roughly 5 times higher than for students who tended to strongly agree that school was relevant to their future goals, whereas the same comparison for PSL was not as strong. Further, as Figure 3.3 shows, exploration of time-dependent associations suggested even greater contrasts in predictive associations among factors. As time progressed and students either accumulated semesters with low SEI scores or did not, those with greater numbers of cumulative low scores tended to graduate at much lower rates than their peers. Based on these preliminary graphical explorations, FSL and FGA were considered the only two factors to have the potential to stand up to the rigorous incremental validity tests, and these findings informed the model building stages of the incremental validity analyses.

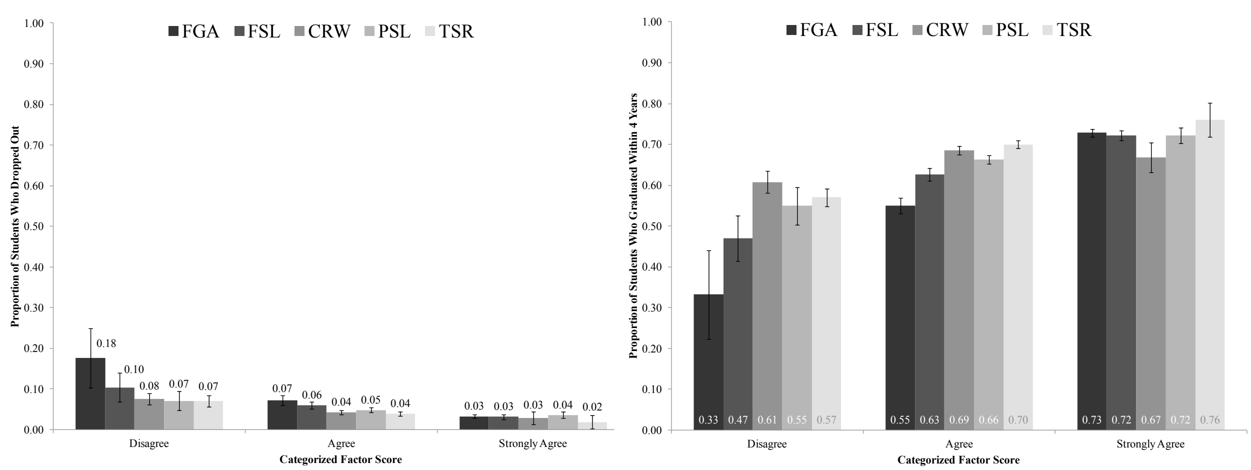

Figure 3.2: Rates for dropout and on-time graduation by 9th grade SEI factor score. Categorized into tended to *Disagree* (< 2.5), tended to *Agree* (2.5-3.4), and tended to *Strongly Agree* (>= 3.5). Error bands represent 95% confidence intervals for the proportion.

.

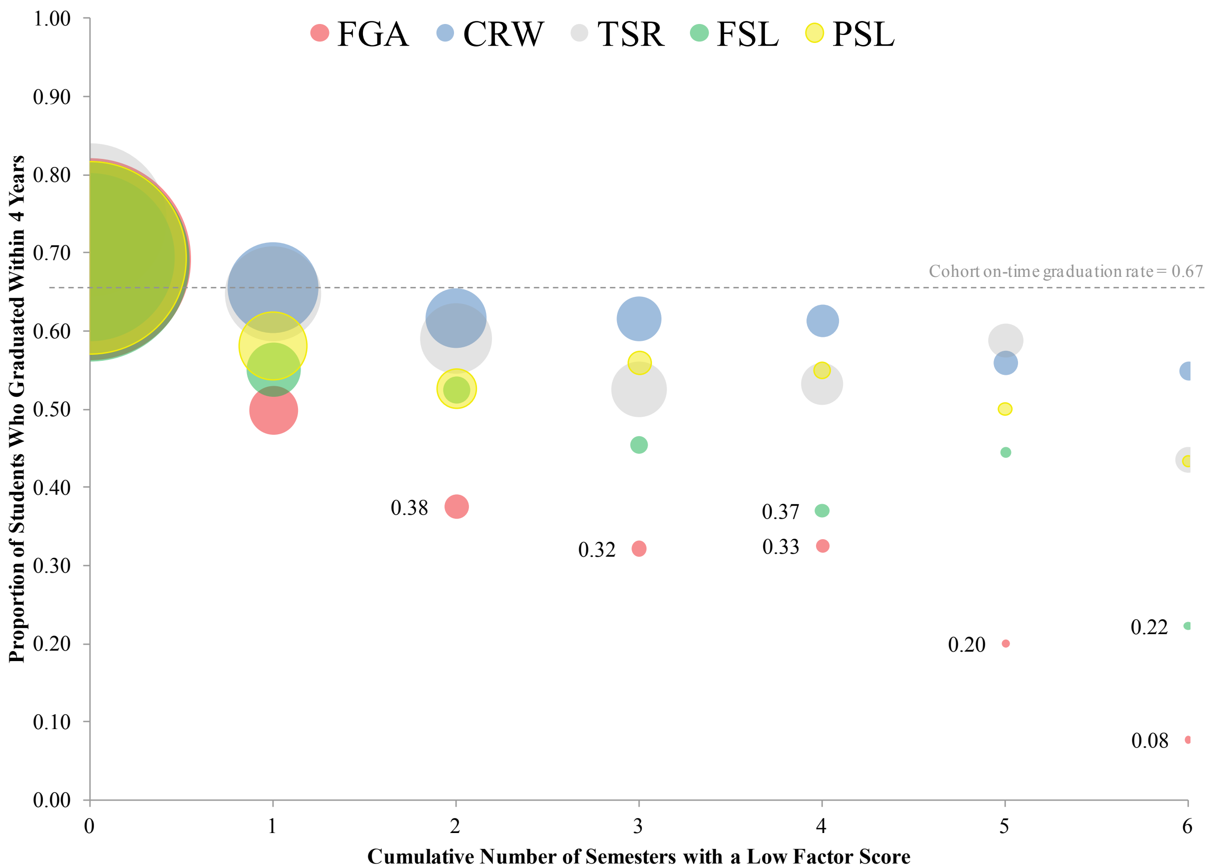

Figure 3.3: Graduation rate by cumulative instances of low SEI scores. On-time graduation rate as a function of each student's cumulative number of semesters in which they reported low levels of engagement. Scores below 2.5, signifying a tendency to respond Disagree or Strongly Disagree across items, were categorized as low. Circle size represents relative size in number of students, ranging from a minimum of 13 (6 semesters with a low score on FGA) to a maximum of 9,262 (0 semesters with a low score on FGA). Data labels shown for rates that are at least 75% lower than the cohort average.

PPV and Sensitivity of Grade Nine Predictors

Figures 3.4 and 3.5 report the results of the dual-criterion tests of PPV and sensitivity for grade nine indicators of dropout and on-time graduation. In these figures, coordinates falling in the shaded areas passed both criteria. In Figure 3.4, the shaded area represents at least a doubling of the dropout rate while also identifying at least 5% of all dropouts. In this graph, the closer a point is to the upper right hand corner the better its power and yield for identifying students who dropped out within four years of their freshmen year. In Figure 3.5, the shaded area represents an on-time graduation rate that was half the cohort average while identifying at least 5% of all students who did not graduate on time. In this graph, the closer a point is to the lower right hand corner the better its power and yield for identifying non-graduates, in other words, students who, for whatever reason, failed to graduate within four years. Overall, behavioral indicators, particularly low attendance and out-of-school suspension, were the most efficient predictors. Students attending less than 90% of enrolled days, for example, were 4 times more likely than the average student to drop out and this indicator correctly identified roughly 35% of all dropouts. Further, only 20% of students with low attendance graduated on time and this indicator identified over 20% of non-graduates. Although less sensitive than behavioral variables, being over-age for grade level stood out as another powerful indicator, with over 20% dropping out over 75% failing to graduate on time. For on-time graduation, low prior achievement was particularly predictive of failing to graduate within four years. Regarding SEI factors, FGA and FSL met both criteria, albeit less substantially than the previously described predictors. Neither of these factors met both criteria for predicting on-time graduation, although FGA was just on the threshold for sensitivity. No other SEI factors met both criteria.

.

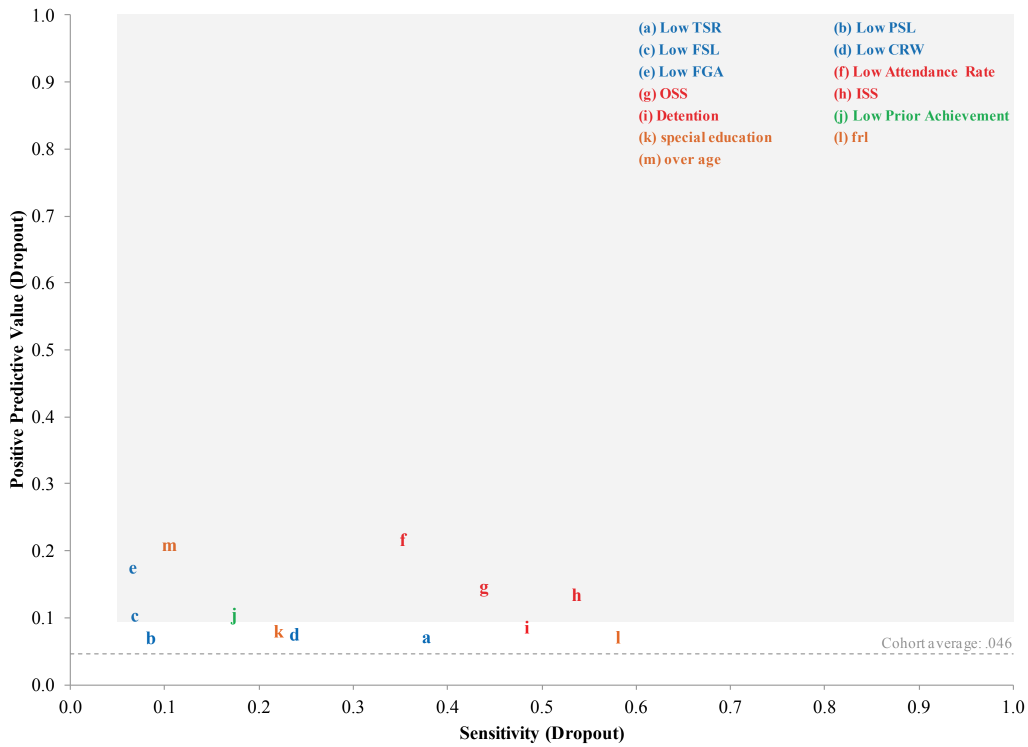

Figure 3.4: Positive predictive value and sensitivity for dropout.

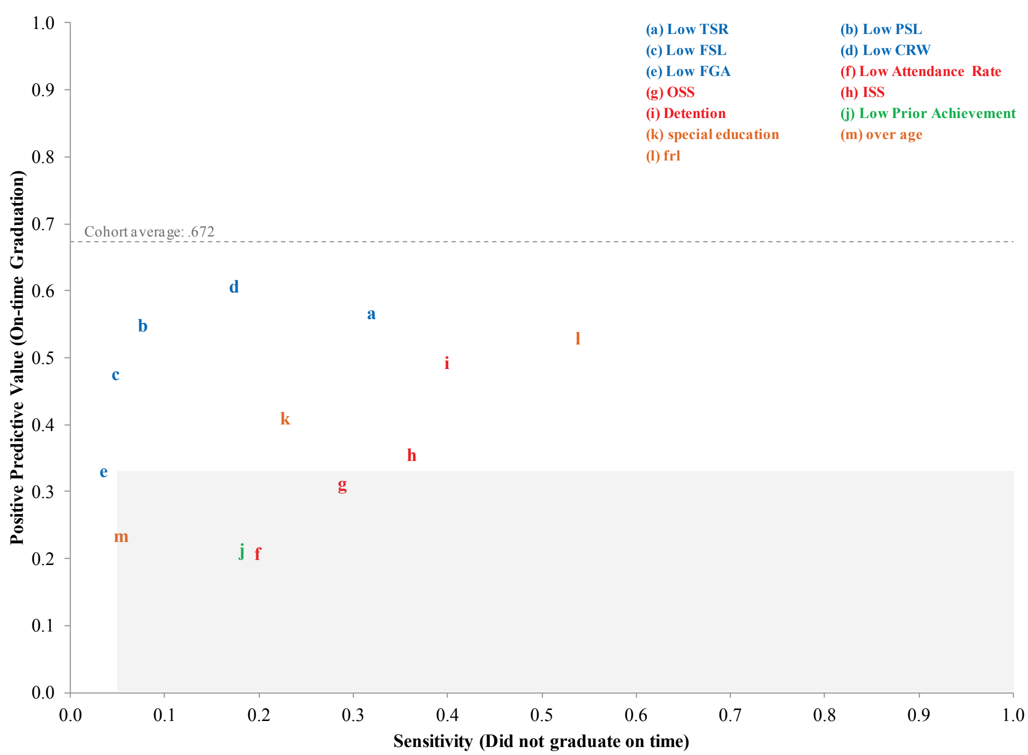

Figure 3.5: Positive predictive value and sensitivity for on-time graduation.

.

Multi-Level Logistic Regression

To keep model building, interpretation, and results as elegant as possible, only the SEI factors meeting the predictive efficiency criteria above (FGA, FSL) were included in subsequent analyses. Results for the estimated multilevel models are summarized in Tables 3.2 and 3.3, including the estimated change in pseudo-R^2 when SEI factors were added, alone and in combination with each other.

Model 1: Unconditional model. In each model, the dichotomous outcome y_{ij} for student i in school j was modeled with a multilevel logistic regression model with a random intercept for schools. In the notation used by Raudenbush and Bryk (2002) and Rabe-Hesketh and Skrondal (2012), the probability that the response was equal to 1 was modeled using a logit link function with the traditional assumption that y_{ij} has a Bernoulli distribution:

$$\text{logit}(\varphi_{ij}) = \eta_{ij}, \quad y_{ij} \backsim \text{Bernoulli}(\varphi_{ij})$$

and the two-level model for Model 1 was:

$$\eta_{ij} = \beta_{0j} \qquad \text{Level 1}$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j} \qquad \text{Level 2}$$

where γ_{00} is the fixed intercept, or the average log-odds of dropout/on-time graduation across schools, while μ_{0j} and τ_{00} represent the random effects, which were assumed to be normally distributed, independent, identically distributed across schools, and independent of covariates (added in later models). Here the important question was whether multilevel modeling was needed. Figure 3.6, displaying the results of a graphical exploration of the data, shows considerable observed variation in dropoout and graduation rates according to school membership in 9th grade (plotted with contrived school ids across the y-axis). Model 1 results indicated that this level-two variance in the response variable was statistically significant for

.

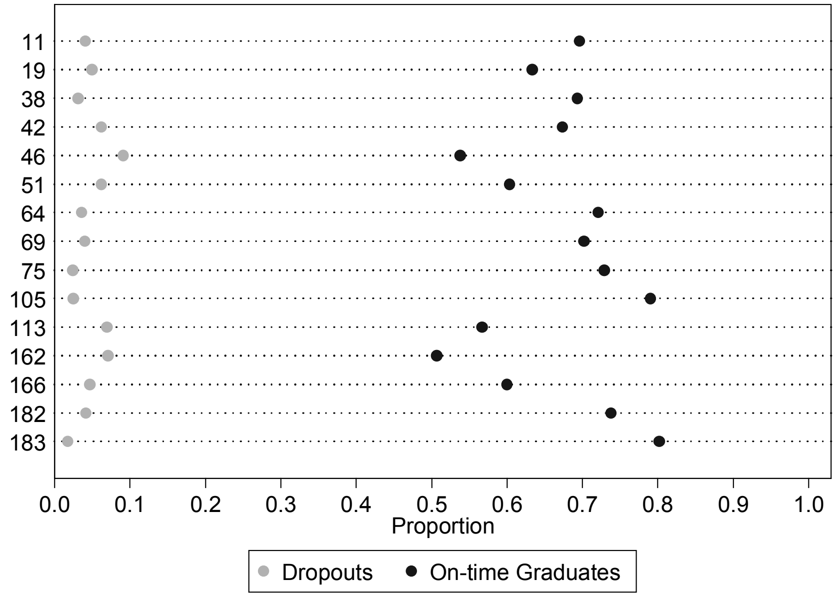

Figure 3.6: Outcome rates by school.

both dropout (ψ = 0.15, 95% CI = 0.05–0.44) and on-time graduation (ψ = .15, 95% CI= .07–.31). Substituting these level-two variance estimates into a conditional ICC equation ($\rho = \frac{\psi}{\psi + \pi^2/3}$; Rabe-Hesketh & Skrondal, 2012) indicated that schools accounted for about 4% of the dropout and graduation variance. Although a variety of studies have found that student-level characteristics commonly account for the majority of the variance in outcomes like dropout (Rumberger & Palardy, 2003), the school-level effects found here are on the extreme low end of estimates reported in the literature. Although the estimated school-level effect was small, the estimated design effect D_{eff}—a numerical representation of the effect of independence violations on standard error estimates (Peugh, 2010)—was large. At 26.8 for these data ($D_{eff} = 1 + (n_c - 1)ICC = 1 + \left(\frac{10{,}067}{15} - 1\right).04$), the D_{eff} suggested the need for a multilevel approach, but the low overall level-two variance suggested that adding school-level variables would not result in better explanatory models.

.

Models 2 to 4: Progression of conditional models. In Models 2 to 4, nine demographic covariates, followed by an achievement covariate, and then by two behavior covariates, were sequentially added. Tables 3.2 and 3.3 show the estimated conditional odds ratio for the outcome of each covariate in these models when FGA and FSL were also included. A pseudo-R^2 statistic without FGA or FSL is presented at the bottom of these tables, followed by the incremental difference in pseudo-R^2 when these variables were added.

Many status variables known to be differentially associated with risk for dropout (e.g., ethnicity, English language learner) were not statistically associated with dropout or on-time graduation when achievement and behavior indicators were added to the model. An exception to this finding was that receiving special education services was estimated to lower the odds of dropout when controlling for achievement, behavior, and cognitive affective engagement. A few demographic variables remained statistically significant throughout all or most models of dropout and graduation. Controlling for all other covariates in the model, being female was estimated to halve the odds of dropout and raise the odds of on-time graduation by 38%, whereas being economically disadvantaged raised the odds of dropout by 52% and lowered the odds of on-time graduation by 30%. The most substantial effect among the demographic variables was for over-age students, whose estimated conditional odds of dropout were 3.8 times higher than age-typical peers, whereas the estimated odds of on-time graduation were 2.1 times lower.

Prior achievement and 9th grade indicators of behavioral engagement and disengagement were also found to be independent and substantially predictive of dropout and on-time graduation. All other variables being equal, a one standard deviation increase in academic achievement was associated with a 2.0 times reduction in the estimated odds of dropping out and a 2.6 times increase in the odds of graduating in four years. Likewise, higher rates of attendance

.

were associated with lower odds of dropout and higher odds of on-time graduation. Regarding signs of behavioral disengagement, having at least one out-of-school suspension at any time in 9th grade was associated with an 89% increase in the odds of dropout and nearly a 2 times decrease in the odds of on-time graduation.

Incremental validity of FGA and FSL. The odds ratio of the effect of FGA and FSL on dropout and on-time graduation were not always statistically significant across models when both factors were included in the model (as shown in Tables 3.2 and 3.3), but each factor was significant in all sub-models when the other factor was excluded. In these sub-models, odds ratios for FGA ranged from 0.34 to 0.43 in models of dropout ($p < .001$ in all cases), indicating that, all other variables being equal, a 1-point decrease in FGA was associated with a 2.3 to 2.9 times increase in the odds of dropping out. For models of on-time graduation, odds ratios ranged from 1.8 to 2.6. For FSL, odds ratios for FSL were all 0.4 in models for dropout and ranged from 1.5 to 1.7 for models for on-time graduation ($p < .001$ in all cases). This evidence in combination with the pseudo-R^2 results suggested that it is likely that FGA and FSL were often explaining much of the same variance in the outcome, with FGA tending to add a little more than FSL to overall variance explained. This would indicate that each of these SEI factor scores showed statistically significant incremental validity over other powerful indicators, but not necessarily over each other. Turning attention more closely to the pseudo-R^2 results, SEI factor scores were shown to add much more to overall variance explained by the model in predictions of dropout than for on-time graduation. In all instances of dropout modeling, these factors—when examined alone or in combination—added enough variance explained to meet Hunsley and Meyer's (2003) criterion. On-time graduation models involving only demographic data met this criterion, but fell just short of it when achievement was added and ≥ 1.5% below the criterion with behavior data.

.

Table 3.3

Dropout multilevel model summaries

Parameters	Model 1: Unconditional	Model 2: Demographic	Model 3: Achievement	Model 4: Behavior
Fixed effects				
Intercept (γ_{00})	0.05 (0.12)	0.04 (0.16)	5.33 (0.70)***	1.29 (0.266)
FGA		0.48 (0.10)***	0.54 (0.11)**	0.64 (0.15)
FSL		0.53 (0.11)**	0.50 (0.10)**	0.48 (0.11)**
Female		0.63 (0.10)**	0.61 (0.09)**	0.49 (0.08)***
Black		1.02 (0.23)	0.85 (0.19)	0.77 (0.19)
Hispanic		0.79 (0.29)	0.67 (0.25)	0.73 (0.30)
Other		0.76 (0.19)	0.70 (0.17)	0.80 (0.21)
ELL		1.40 (0.53)	1.20 (0.46)	0.98 (0.41)
FRL		1.96 (0.34)***	1.74 (0.31)**	1.52 (0.30)*
Over-age		5.32 (2.28)***	4.20 (1.84)**	3.82 (1.73)**
Special Ed.		1.28 (0.26)	0.69 (0.15)	0.61 (0.15)*
Gifted		0.27 (0.08)***	0.55 (0.18)	0.66 (0.23)
Avg. Achievement			0.45 (0.06)***	0.49 (0.07)***
Percent Attendance				0.86 (0.02) ***
OSS				1.89 (0.43)**
Variance components				
Intercept (ψ)	.150 (.082)	.045 (.056)	.000 (.015)	.014 (.001)
Conditional ICC (ρ)	.043 (.024)	.013 (.017)	.000 (.005)	.000 (.004)
Pseudo-R^2				
No FGA or FSL	.051	.131	.183	.334
Incremental Difference in Pseudo-R^2				
with FGA		+.057	+.045	+.024
with FSL		+.050	+.045	+.031
with FGA & FSL		+.070	+.058	+.036

Note. Balanced case-control sample (n=930). Prior correction to level-1 intercept (King & Zeng, 2001). Dropout does not include students who students who left for unknown reasons or students who earned a certificate of completion or a special education diploma.
*** $p < .001$ **$p < .01$ *$p < .05$

.

Table 3.4

On-time graduation multilevel model summaries

Parameters	Model 1: Unconditional	Model 2: Demographic	Model 3: Achievement	Model 4: Behavior
Fixed effects				
Intercept (γ_{00})	2.05 (.21)***	2.54 (0.17)***	1.65 (0.11)***	1.82 (0.12)***
FGA		2.42 (0.19)***	2.05 (0.16)***	1.72 (0.14)***
FSL		1.12 (0.07)	1.20 (0.09)*	1.13 (0.08)
Female		1.26 (0.06)***	1.24 (0.06)***	1.38 (0.07)***
Black		0.89 (0.06)	1.15 (0.08)	1.02 (0.08)
Hispanic		0.84 (0.09)	0.96 (0.11)	0.87 (0.10)
Other		1.14 (0.09)	1.34 (0.11)***	1.03 (0.09)
ELL		0.78 (0.09)*	1.05 (0.12)	1.06 (0.13)
FRL		0.58 (0.03)***	0.68 (0.04)***	0.77 (0.05)***
Over-age		0.27 (0.04)***	0.37 (0.06)***	0.48 (0.09)***
Special Ed.		0.40 (0.03)***	0.89 (0.07)	0.94 (0.08)
Gifted		4.13 (0.37)***	1.67 (0.16)***	1.53 (0.16)***
Achievement			2.71 (0.12)***	2.57 (0.11)***
Percent Attendance				1.18 (0.01)***
OSS				0.53 (0.04)***
Variance components				
Intercept (ψ)	.147 (.053)	.024 (.013)	.018 (.010)	.014 (.008)
Conditional ICC (ρ)	.042 (.017)	.007 (.004)	.005 (.003)	.004 (.003)
Pseudo-R^2				
No FGA or FSL	.033	.146	.218	.312
Incremental Difference in Pseudo-R^2				
with FGA		+.033	+.022	+.010
with FSL		+.012	+.010	+.004
with FGA & FSL		+.033	+.022	+.011

n=10,067

***$p < .001$ **$p < .01$ *$p < .05$

.

Discussion

The negative impact of dropout on individuals and society underscores the importance of efforts to raise school completion rates. One necessary component of systematic efforts is effective early warning screening for students at risk (Balfanz et al., 2007; Christenson & Thurlow, 2004). No single variable predicts dropout well enough to do the job on its own (Gleason & Dynarski, 2002), but identification of students falling off the graduation path improves when predictive models include a variety of efficient indicators (Rumberger & Lim, 2008; Gleason & Dynarski, 2002; Balfanz et al., 2007). Student engagement is a multifaceted construct, consisting of at least behavioral, cognitive, and affective features (Fredricks et al., 2004), and the substantial predictive value of behavioral engagement indicators (e.g., attendance, discipline) is well known (Rumberger & Lim, 2008). The practical advantages of measuring cognitive and affective engagement, however—psychological constructs that require psychometric instruments to be measured reliably—is less well known. The purpose here was to investigate the potential of including such measures in risk detection efforts, by testing the associations of cognitive and affective engagement in the 9th grade with 4 year outcomes.

The focus of this study was on examining the predictive efficiency and incremental validity of the SEI, a measure of cognitive and affective engagement, for early identification of students falling off the graduation path. In this study, alongside a variety of data commonly available in school records, the SEI was subjected to tests of its sensitivity and positive predictive value as an individual indicator of dropout and on-time graduation. Although not nearly as efficient in predicting dropout or graduation as a behavioral indicator like low attendance rate, two factors on the SEI, FGA and FSL, demonstrated considerable efficiency for a self-report measure in identifying students who were known to drop out within four years of their freshman semester. Next, the predictive utility of these SEI factors was further assessed by examining their incremental validity over commonly available school data. This was accomplished by studying their unique statistical contribution to the variance explained in dropout and on-time graduation through a multilevel logistic regression model that controlled for a variety of school data, including several well-known strong predictors of high school outcomes. Here the SEI showed less promise as a predictor of on-time graduation, but in stringent incremental validity tests it performed considerably well as a predictor of dropout by remaining statistically significant and contributing unique explained variance to models that already included 12 other variables, several of which—like prior achievement and student age—were shown to be highly and independently predictive of dropout.

Dropout and graduation rates are known to differ considerably across various demographic characteristics, but research has shown that such effects often depend on which other variables are included in the study (Rumberger & Lim, 2008). In the multilevel analysis here, the effects on the odds of dropping out and on graduating on time of ethnicity and English-language-learner status—two demographic characteristics well-known to associate differentially with these outcomes (Rumberger & Lim, 2008)—were null once achievement, behavior, and engagement indicators were included in the analysis. Similarly, according to the full model, receiving special education services in 9^{th} grade was shown to statistically reduce the odds of dropping out, and showed no effect on the odds of whether or not a student graduated in 4 years. Such findings are common in the abundance of literature on this topic, and as Rumberger and Lim (2008) put it, they indicate that, in the case of ethnicity and special education status, “the observed relationship . . . can often be explained by other factors.” The insignificant effects found here for language status should be interpreted with caution because the data did not allow for English proficiency to be controlled for, only whether or not a student spoke English as a native language. Other studies (e.g., Griffin & Heidorn, 1996; Perreira, Harris, & Lee, 2006; Lutz, 2007) have found protective benefits of language proficiency when controlling for other background and academic characteristics.

The higher graduation rates and lower dropout rates found here for females, however, are less representative of findings in the literature. Many studies have found females to have lower dropout rates and higher graduation rates than males, but in general the opposite trend has been found when attitudes, behaviors, and achievement are taken into account (Rumberger & Lim, 2008). The positive effects found here may be due to characteristics of the district studied—a large suburban school system near a metropolitan city—as some studies have found similar results when looking at gender effects between various subpopulations. For example, similar to findings here that being female is a protective factor, Lichter, Cornwell, and Eggebeen (1993) found lower dropout rates for females than for males among students in central cities and suburbs, but the reverse trend in rural areas.

All other variables in the final model that were found to show independent predictive effects when controlling for a variety of other data were consistent with the literature. Free-/reduced-price lunch eligibility, for example, is often used as a proxy for low family financial resources, which is commonly associated with less means and opportunities for enriched learning outside of school. According to Rumberger and Lim (2008), most studies have shown that students from lower-income households are more likely to dropout and less likely to graduate, as was found here, even when controlling for all other variables in the model. Being over-age for grade level was also found to substantially impact the odds of dropout and graduation. Most studies report similar risk increasing effects for students greater than 1 to 2 years older than their grade-level peers (Rumberger & Lim, 2008). Also found here was that higher achievement scores in 8th grade and higher rates of attendance in 9th grade were each independently predictive of a lower likelihood of dropping out and a higher likelihood of graduating on time; whereas receiving at least one out-of-school suspension (OSS) in 9th grade greatly reversed the effect. OSS, an indicator of school misbehavior, may be viewed as a behavioral marker of disengagement, which many studies have found to be positively associated with dropout and negatively associated with graduation even when prior academic achievement and family background were taken into account (Rumberger & Lim, 2008).

In all, the findings from the above analyses fit well with the body of research on the multivariate effects of risk and protective factors on dropout and graduation. But the main focus here was on the SEI and its predictive qualities. Prior research has investigated the internal consistency, latent factor structure, and measurement invariance of the SEI (Appleton et al., 2006; Betts et al., 2010). This paper adds to this growing body of research on the SEI's construct validity through its examination of the instrument's long-term predictive validity for high school graduation and dropout, addressing questions raised in the original validation study (Appleton et al., 2006) about the relationship of the SEI with relevant educational variables. Further, by following a freshman cohort for four years and linking their 9th grade data to their individual educational outcomes, the analyses provide an indication of what can be expected when the SEI is used in an applied setting, which should be valuable to stakeholders interested in meaningful evidence relevant to screening for risk of school failure.

Similar to the pattern of small correlations reported by Appleton et al. (2006) for SEI factor associations with some relevant outcomes (like grade point average, standardized tests, and suspensions), findings in this study found that some SEI factors did not appear to be meaningfully predictive of on-time graduation, at least in terms of scores representative of a student's self-perceptions of engagement at a single moment in time in 9th grade. Unlike Appleton et al.'s early findings, however, this study did find clear and considerable links between student responses on the FGA and FSL factors and later outcomes. Further, these scores represented the summary of roughly 10 self-report items completed by students at a time in the early days of their high school career—not data from a time- and labor-intensive research initiative. With this in mind, as Gwinnett County Public Schools continues with its engagement and advisement initiative and more data becomes available, the next step in understanding the practical predictive potential of the SEI might be to examine what early SEI *trajectories* explain about the likelihood of dropping out. Janosz, Archambault, Morizot, and Pagani (2008) found very strong relationships between engagement trajectories and dropout, yet engagement in the context of their analyses was largely composed of behavioral measures, and no study was given to incremental differences in explanatory power. It will be important to investigate whether understanding factors like FGA or FSL in developmental terms add value to prediction. If so, it is possible that the incremental validity found in the present study for SEI factors in explaining dropout may be even more pronounced in the early middle school years and when considering trajectories rather than single points in time.

The findings in this study support findings from over 25 years of dropout research that outward indicators of disengagement from school, such as disciplinary problems or low attendance rates, are powerful predictors of dropping out (Rumberger & Lim, 2008). It is possible that for many students, however, that disengagement in the middle school years starts with a notion not yet expressed in such obvious terms. An unstable path of FGA scores, for instance, or perhaps an unusually rapid drop in FGA scores from 6th to 7th grades may predate full-blown behavioral disengagement for these students. Further, while many students may manifest disengagement through disciplinary problems and low attendance, not everyone who dropped out showed these early signs. Early changes in FGA or FSL may help to better identify and better understand students who fall off the graduation path. Investigating these possibilities is important to understanding the full potential of the measure and for a fuller theoretical understanding of the developmental aspects of engagement.

An important limitation of this study was the retrospective, observational nature of the methods employed in it, which preclude statements of causal relationships among any of the variables in our analyses. Further, although our sample was large and diverse, it represented one district in the country; further generalizations to non-representative populations may be unfounded. Others are encouraged to replicate these analyses in other districts and across districts, particularly in rural settings or in settings with substantial variability in risk across schools for dropout. The time constraints of our data should also be noted, because we were limited to studying on-time graduation only. Of the 8.1% of students in the cohort who were still enrolled in Fall 2011, many may have graduated later, and other non-graduates may have re-enrolled at a later date or graduated in other districts.

Overall, in combination with other research on the SEI, the findings of this study suggest that greater confidence may be placed in the interpretation of SEI scores for students in the process of disengaging from school and with long-term educational objectives in mind. Although many previously reported correlations between SEI scores and indicators of behavioral engagement/disengagement and achievement were small to non-significant, the results of this study show that—when following students over several years there are significant, predictable, and educationally meaningful associations between some SEI scores and relevant outcomes.

Early screening for risk is a necessary component of systematic, evidence-based prevention programs (Christenson & Thurlow, 2004; Jimerson, Reschly, & Hess, 2008), and while the findings here in no way suggest that the SEI should be used on its own as a reliable, robust screener for educational risk, they do indicate promising potential for the SEI's inclusion in multifactor risk identification efforts. While student engagement is generally regarded as a multifaceted construct (Fredricks et al., 2004), most engagement studies have involved primarily, or only, behavioral indicators. Data on cognitive and affective engagement may be more challenging to gather than behavioral engagement indicators, but this study found distinct, additive value when incorporated into a multifactor model.

CHAPTER 4

SUMMARY AND CONCLUSIONS

The primary goal of this two-study dissertation was to better understand the validity of the Student Engagement Instrument (SEI) by investigating its longitudinal characteristics. To accomplish this, two related studies were conducted, one on how students' SEI scores change through adolescence, the other on how well 9th grade SEI responses predict dropout and graduation. Using the responses of over 40,000 students, the first study used an accelerated design to examine several longitudinal aspects of the SEI. The chief objective of the first study was to get a comprehensive view of how SEI scores change through the middle and high school grades, exploring the annual rank-order stability of the instrument, the normative change of SEI scores semester to semester, and whether there was also reliable evidence of non-normative change. The aim of the second study was to explore the practical predictive utility of the SEI by estimating its predictive efficiency and incremental validity beyond school record data. A cohort of roughly 10,000 first-time ninth graders was followed for four years, and the associations of their 9th grade SEI scores with dropping out and with graduating on time were assessed in terms of positive predictive value, sensitivity, and ability to explain variance in outcomes when accounting for powerful predictors commonly available in school records.

Findings from the first study indicated moderate annual retest stability for all SEI factor scores, results that are in line with expectations based on theory (Conley, 1984). Regarding evidence of normative and non-normative change, analyses of mean-level change showed gradual decreases over time for each factor score, and analyses of reliable change suggested that non-normative trajectories are likely to exist. These results fit well within the context of similar empirical research (Janosz, Archambault, Morizot, & Pagani, 2008; Wylie & Hodgen, 2012). In the second study, results of the sensitivity and positive predictive value analyses indicated that, while not nearly as efficient as a predictor like low attendance, low scores on the SEI adequately identified dropouts in terms of a dual-criterion test of predictive power and yield. Finally, in a multivariate analysis involving a wide variety of powerful predictors of dropout and graduation, lower scores on the SEI were shown to increase the odds of dropping out. Further, SEI factors contributed a reasonable amount of explanatory power to the model when predicting dropout.

The common thread running through these findings is that, by showing stability and change in line with theory as well as predictive efficiency and incremental validity, the SEI shows promise as a longitudinal measure. But more research is needed before its full value can be understood. One of the primary implications of the rank-order stability findings is that, in light of the moderate retest stability of the SEI from one year to the next, developmental trends in a student's scores may provide richer, more predictive information than a single point in time about students' engagement in school. When enough data become available, it would be particularly valuable to explore non-normative trajectories in the middle school years, and if these exist, whether students within certain pathways are much more likely than others to drop out. As pointed out in Chapter 3, not all students who drop out show outward indicators early on. It is possible that disengagement in the middle school years starts with much subtler signs for many students, such as thoughts that school is becoming less relevant to future goals. Investigating these possibilities is important to understanding not only the full potential of the SEI for use in both research and practice but also for being able to achieve a fuller theoretical understanding of the nature of engagement and how it develops through the school years.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological tests*. Washington, DC: American Educational Research Assocation.

Appleton, J. J. (2012). Systems consultation: Developing the assessment-to-intervention link with the Student Engagement Instrument. In S.L. Christenson, A.L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 725–741). New York: Springer. doi: 10.1007/978-1-4614-2018-7_35

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,* 369-386. doi: 10.1002/pits.20303

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427-445. doi: 10.1016/j.jsp.2006.04.002

Appleton, J. J., Reschly, A. L., & Martin, C. (2013). *Research to practice: Measuring and reporting student engagement. Manuscript under review*.

Asendorpf, J. B. (1992). Continuity and stability of personality traits and personality patterns. In J. B. Asendorpf & J. Valsiner (Eds.), *Stability and change in development: A study of methodological reasoning* (pp. 116–142). Newbury Park, CA: SAGE.

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., Tahan, K. (2011). *The condition of education 2011* (NCES 2011-033). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223-235. doi: 10.1080/00461520701621079

Betts, J., 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,* 84-93. doi: 10.1037/a0020259

Blonigen, D. M., Hicks, B. M., Krueger, R. F., Patrick, C. J., & Iacono, W. G. (2006). Continuity and change in psychopathic traits as measured via normal-range personality: A longitudinal-biometric study. *Journal of Abnormal Psychology*, *115*, 85-95. doi: 10.1037/0021-843X.115.1.85

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*, *27*(2), 61-73. doi: 10.1037/a0029229

Caruso, J. C., & Cliff, N. (1997). Empirical size, coverage, and power of confidence intervals for Spearman's rho. *Educational and Psychological Measurement*, *57*(4), 637-654. doi: 10.1177/0013164497057004009

Chapman, C., Laird, J., and KewalRamani, A. (2010). *Trends in High School Dropout and Completion Rates in the United States: 1972–2008* (NCES 2011-012). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Christenson, S. L., Reschly, A. L., Appleton, J. J., Berman-Young, S., Spanjers, D. M., & Varro, P. (2008). Best practices in fostering student engagement. In A. Thomas & J. Grimes (Eds.) *Best practices in school psychology* (5th ed., pp. 1099–1119). Bethesda, MD: National Association of School Psychologists.

Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). Preface. In S.L. Christenson, A.L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. v–ix). New York: Springer. doi: 10.1007/978-1-4614-2018-7

Christenson, S. L., & Thurlow, M. L. (2004). School dropouts prevention considerations, interventions, and challenges. *Current Directions in Psychological Science*, *13*(1), 36-39. doi: 10.1111/j.0963-7214.2004.01301010.x

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. doi: 10.1037/0033-2909.112.1.155

Conley, J. J. (1984). The hierarchy of consistency: A review and model of longitudinal findings on adult individual differences in intelligence, personality and self-opinion. *Personality and Individual Differences*, *5*(1), 11-25. doi: 10.1016/0191-8869(84)90133-8

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods*, *1*(3), 236-248. doi: 10.1037/1082-989X.1.3.236

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59,* 117-142.

Finn, J. D. (2006). *The adult lives of at-risk students: The roles of attainment and engagement* *in high school.* (NCES 2006-328). Washington, DC: National Center for Education Statistics, U.S. Department of Education.

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. doi: 10.3102/00346543074001059

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: A description of 21 instruments.* (Issues & Answers Report, REL 2011-No. 098). Washington, DC: U.S. Department of Education, Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

Gardner, D. P. (1983). *A nation at risk.* Washington, DC: U. S. Department of Education.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* Cambridge, MA: Cambridge University Press.

Georgia Department of Education (2013). *Student enrollment by grade level.* [Online data table]. Retrieved from http://app.doe.k12.ga.us/fte.html

Gleason, P., & Dynarsky, M. (2002). Do we know whom to serve? Issues in using risk factors to identify dropouts. *Journal of Education for Students Placed at Risk*, 7(1), 25-41. doi: 10.1207/S15327671ESPR0701_3

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, *93*(1), 3-13. doi: 10.1037/0022-0663.93.1.3

Griffin, B. W., & Heidorn, M. H. (1996). An examination of the relationship between minimum competency test performance and dropping out of school. *Educational Evaluation and Policy Analysis, 18*(3)*,* 243–252.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, *65*(2), 202-226. doi: 10.1177/0013164404267287

Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*, *15*(4), 446-455. doi: 10.1037/1040-3590.15.4.446

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12-19. doi: 10.1037/0022-006x.59.1.12

Janosz, M. (2012). Outcomes of engagement and engagement as an outcome: Some consensus, divergences, and unanswered questions. In S.L. Christenson, A.L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 695-703). New York: Springer. doi: 10.1007/978-1-4614-2018-7_33

Janosz, M., Archambault, I., Morizot, J., & Pagani, L. S. (2008). School engagement trajectories and their differential predictive relations to dropout. *Journal of Social Issues, 64*(1), 21-40. doi: 10.1111/j.1540-4560.2008.00546.x

Jimerson, S. R., Reschly, A. L., & Hess, R. S. (2008). Best practices in increasing the likelihood of high school completion. In A. Thomas & J. Grimes (Eds.), *Best practices in school* *psychology* (5th ed., pp. 1085-1097). Bethesda, MD: National Association of School Psychologists.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political* Analysis, *9*(2), 137–163. doi: 10.1093/oxfordjournals.pan.a004868

Lacy, M. G. (1997). Efficiently studying rare events: Case-control methods for sociologists. *Sociological Perspectives*, *40*(1), 129-154. doi: 10.2307/1389496

Legters, N. E., McPartland, J. M., & Balfanz, R. (2004). Scaling up Talent Development High Schools: Lessons learned from comprehensive high school reform. In T. K. Glennan, S. J. Bodilly, J. R. Galegher, & K. A. Kerr (Eds.), *Expanding the reach of education reforms: Perspectives from leaders in the scale-up of educational interventions* (pp. 379–431). Santa Monica, CA: RAND Corporation.

Lehr, C. A., Johnson, D. R., Bremer, C. D., Cosio, A., & Thompson, M. (2004). *Essential tools: Increasing rates of school completion: Moving from policy and research to practice.* Minneapolis, MN: University of Minnesota, Institute on Community Integration, National Center on Secondary Education and Transition.

Lewis, A. D., Huebner, E. S., Reschly, A. L., & Valois, R. F. (2009). The incremental validity of positive emotions in predicting school functioning. *Journal of Psychoeducational Assessment*, *27*(5), 397-408. doi: 10.1177/0734282908330571

Lichter, D. T., Cornwell, G. T., & Eggebeen, D. J. (1993). Harvesting human-capital: Family structure and education among rural youth. *Rural Sociology*, *58*(1)*,* 53–75. doi: 10.1111/j.1549-0831.1993.tb00482.x

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Lovelace, M. D. (2010). *A construct validity study of the student engagement instrument: Fit with at-risk youth and degree of association with behavioral indicators of disengagement* (Master's thesis). Retrieved from http://www.libs.uga.edu/etd/

Lovelace, M. D. (2011, February). The SEI and the behaviorally disengaged: Estimates of model fit with at-risk youth and relative prevalence of outcomes among students with low, medium and high SEI scores. In A. L. Reschly (Chair), *Advances in Student Engagement and Preparation for College*. Symposium conducted at the Annual Convention of the National Association of School Psychologists, San Francisco, CA.

Lutz, A. (2007). Barriers to high school completion among immigrant and later-generation Latinos in the USA: Language, ethnicity and socioeconomic status. *Ethnicities*, *7*(3), 323–342. doi: 10.1177/1468796807080232

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American educational research journal*, *37*(1), 153-184. doi: 10.3102/00028312037001153

Martin, A. J. (2003). The Student Motivation Scale: Further testing of an instrument that measures school students' motivation. *Australian Journal of Education*, *47*(1), 88-106. doi: 10.1177/000494410304700107

Miyazaki, Y., & Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods*, *5*(1), 44-63. doi: 10.1037/1082-989X.5.1.44

Moreira, P. A., Vaz, F. M., Dias, P. C., & Petracchi, P. (2009). Psychometric properties of the Portuguese version of the Student Engagement Instrument. *Canadian Journal of School Psychology*, *24*(4), 303-317. doi: 10.1177/0829573509346680

National Research Council and the Institute of Medicine. (2004). *Engaging schools: Fostering high school students' motivation to learn*. Washington, DC: The National Academies Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Otis, N., Grouzet, F. M., & Pelletier, L. G. (2005). Latent motivational change in an academic setting: A 3-year longitudinal study. *Journal of Educational Psychology*, *97*(2), 170-183. doi: 10.1037/0022-0663.97.2.170

Perreira, K. M., Harris, K. M., & Lee, D. (2006). Making it in America: High school completion by immigrant and native youth. *Demography*, *43*(3), 511-536. doi: 10.1353/dem.2006.0026

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*, 85–112. doi: 10.1016/j.jsp.2009.09.002

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed., Vols. 1–2). College Station, TX: StataCorp.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reschly, A. L., Betts, J., & Appleton, J. J. (2013). *An examination of the validity of two measures of student engagement*. Manuscript submitted for publication.

Reschly, A., & Christenson, S. L. (2006).Promoting school completion. In G. Bear & K. Minke (Eds.), *Children's needs III: Understanding and addressing the developmental needs of children*. Bethesda, MD: National Association of School Psychologists.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S.L. Christenson, A.L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 3–19). New York: Springer. doi: 10.1007/978-1-4614-2018-7_1

Reschly, A.L., Huebner, E.S., Appleton, J.J., & Antaramian, S. (2008). Engagement as flourishing: The contribution of positive emotions and coping to adolescents' engagement at school and with learning. *Psychology in the Schools, 45*(5), 419-431. doi: 10.1002/pits.20306

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25. doi: 10.1037/0033-2909.126.1.3

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*(1), 1–25. doi: 10.1037/0033-2909.132.1.1

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69*(4), 617-640. doi: 10.1111/1467-6494.694157

Rouse, C. E. (2005, September). *The labor market consequences of an inadequate education*. Paper prepared for the Equity Symposium on the Social Costs of Inadequate Education, Teachers College, Columbia University. Retrieved July 2011, from http://devweb.tc.columbia.edu/ manager/symposium/Files/77_Rouse_paper.pdf

Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it.* Cambridge, MA: Harvard University Press.

Rumberger, R. W., & Lim, S. A. (2008). Why students drop out of school: A review of 25 years of research. *California Dropout Research Project* (Report #15). Santa Barbara, CA: UC Santa Barbara, Gevitz Graduate School of Education.

Rumberger, R. W., & Palardy, G. J. (2004). Multilevel models for school effectiveness research. In D. W. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for the Social Sciences* (pp. 235–258). Thousand Oaks, CA: Sage.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, *57*(1), 1–10. doi: 10.1037/a0018082

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement*, *23*, 153-158. doi: 10.1177/001316446302300113

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, *85*(4), 571–581. doi: 10.1037/0022-0663.85.4.571

Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S.L. Christenson, A.L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 21–44). New York: Springer. doi: 10.1007/978-1-4614-2018-7_2

Snyder, T. D. (1993). *One hundred twenty years of American education: A statistical portrait.* National Center for Educational Statistics, Washington, D. C.: U.S. Department of Education.

StataCorp (2011). *Stata statistical software: Release 12*. College Station, TX: StataCorp LP.

Stillwell, R., & Sable, J. (2013). *Public school graduates and dropouts from the Common Core of Data: School year 2009-10: First look (Provisional data)* (NCES 2013-309). U.S. Department of Education, Washington, D.C.: National Center for Education Statistics. Retrieved March 6, 2013 from http://nces.ed.gov/pubsearch.

Tonry, M., Ohlin, L. E., & Farrington, D. P. (1991). *Human development and criminal behavior: New ways of advancing knowledge.* New York: Springer-Verlag.

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R., & Davis-Kean, P. (2006). Development of achievement motivation. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology (6th ed.): Vol. 1. Social, emotional, and personality development* (pp. 933–1002). New York: John Wiley & Sons. doi: 10.1002/9780470147658

Wylie, C., & Hodgen, E. (2012). Trajectories and patterns of student engagement: Evidence from a longitudinal study. In S.L. Christenson, A.L. Reschly & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 585–599). New York: Springer. doi: 10.1007/978-1-4614-2018-7_28