THE ASSESSMENT OF STUDENTS' SOCIAL-EMOTIONAL COMPETENCIES AND ACADEMIC ACHIEVEMENT

By

DOMINIC C. MOCERI

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Psychology

written under the direction of

Maurice J. Elias, Ph.D.

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October 2015

ABSTRACT OF THE DISSERTATION

The Assessment of Students' Social-Emotional Competencies and Academic Achievement

By DOMINIC C. MOCERI

Dissertation Director:
Maurice J. Elias, Ph.D.

In order to reduce the burden of mental illness a major shift in intervention research and clinical practice must occur (Kazdin & Blase, 2011). One way to create a major shift in the mental health well-being of youth is through universal school-based prevention and promotion interventions, such as social-emotional learning (SEL). The five core SEL competencies are self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (CASEL, 2005). A recent meta-analysis revealed that universal school-based SEL interventions lead to increases in social emotional skills, positive attitudes towards oneself and others, positive social behaviors, and academic performance, as well as decreases in conduct problems and emotional distress (Durlak et al., 2011). While there is currently no recognized feasible, scalable assessment system for SEL, there is an assessment system in educational practice already utilized in the vast majority of schools: the report card, which contains not only letter grades but also comments on student behavior. This dissertation is the first study to systematically and empirically study how the behavioral ratings and comments sections of the report card relate to SEL, academic grades, and standardized tests in a manner consistent with SEL

theory. Over 1,000 report cards from an ethnically diverse, large high school were chosen as a case study. Ten report card comment categories were created using grounded theory; this included a partial representation of the presence and absence of SEL skills/behaviors as well as non-SEL skills/behaviors. Overall, concurrent analyses revealed that report card comments had a small effect size on attendance and a large effect size on letter grades. Comments partially explained academic disparities for gender and ethnicity for these analyses. These findings applied to both "subjective" (i.e., language) and "objective" (i.e., mathematics) academic areas. Additionally, the effects of demographics, attendance, comments, and letter grades on standardized test scores were examined. Finally, the predictive validity of the comments were examined, after controlling for prior academic success. Implications, limitations and future directions for creating individualized, sustainable, and scalable SEL assessments using report card comments are discussed.

ACKNOWLEDGEMENTS

I would like to begin by acknowledging the students, teachers, and school staff who participated in this study. Without their report cards, this dissertation would not exist. I would like to send a extra thank you to the key school informant who provided me with information about the history of the school report cards at the participating school.

Next, I would like to thank those who helped code the report card comments. First, there were the two intelligent, friendly and productive undergraduate research assistants at the Social-Emotional Learning Lab, who have been helpful in many, many ways on this project: Molly S. Kaufman and Arthi Yerrmilli. Second, there were the two school psychologists with SEL expertise who generously gave their time despite being unaffiliated with the participating school: Barry Barbarasch and Jeffrey Selman.

Next, I would like to thank my committee members for their contributions to my education and dissertation. Dan Fishman provided me with knowledge and encouragement to make sure that my research is useful to practicing clinicians, teachers and other professionals, which influenced the approach I took for my conceptualization. Yakov Epstein pushed me to more carefully consider the context in which report cards are completed, which influenced my data analytic plan. Edward A. Selby helped me through challenging statistical hurdles that occurred when I began my data analyses and aided in the interpretation of my results. Mark T. Greenberg provided the groundwork through which I envision implementation of school-based preventative interventions. And my mentor, Maurice J. Elias. He has provided me with guidance, feedback, and support throughout my entire graduate school career. His willingness to merge my interests in implementation science with his projects has allowed my master's thesis, my dissertation,

and the rest of my research to consistently be near and dear to my heart instead of being items checked off of a list on the way to graduation.

Next, I would like to thank my fellow graduate student lab mates, Jazmin A. Reyes-Portillo, Gwyne White, and Cesalie Stepney, at the Social-Emotional Learning Lab at Rutgers. They have provided numerous hours of theoretical, logistical and emotional support before this dissertation, during this dissertation, and, hopefully, after this dissertation.

Next, I would like to thank my clinical psychology cohort: Alison M. Staples, Amy P. Hansford, Darren Stolow, Irene Zilber, Jazmin A. Reyes-Portillo (again), Jessica Breland and Vivienne Yeh. There are not words to express how privileged I feel to have not been not only colleagues with these seven individuals but also good friends. I could not imagine graduate school without them.

Finally, I would like to thank my family. My sister Anna, my mom, my dad, and my grandparents have been strong supporters of my education and have provided much love and emotional support. And, most importantly, I would like to thank my wife, Christy. Her unwavering love, support, patience, and compassion throughout our relationship, especially during the seemingly never ending intrusion of graduate school, has been nothing short of uncanny and astonishing.

TABLE OF CONTENTS

PAGE

ABSTRACT..... ii

ACKNOWLEDGEMENTS..... iv

LIST OF TABLES..... vii

LIST OF FIGURES..... viii

INTRODUCTION..... 1

METHOD..... 17

RESULTS..... 21

DISCUSSION..... 45

REFERENCES..... 56

APPENDICES

A. CASEL 5: Domains, Definitions, and Example..... 60

B. A Comparison of the Theoretical and and Logic Models..... 65

C. Detailed Analyses of Total Comments and Letter Grades..... 67

D. Repeated measure ANOVAs of Grade Level..... 69

E. Repeated measure ANOVAs of Ethnicity..... 73

F. Repeated measure ANOVAs of Gender..... 76

G. Repeated measure ANOVAs of Lunch Status..... 79

LIST OF TABLES

Table		Page
1.	Participant Demographics and Averages	82
2.	List of 25 Report Card Comments	84
3.	Comment Categories: 3-Way Consensus	85
4.	Comment Categories: 10 Final Categories	87
5.	Frequency of Comment Categories for Language/Literature: All Grades	89
6.	Frequency of General Comment Categories for All Quarters: All Grades	90
7.	Frequency of Positive and Negative Comments for Language/Literature: All Grades	91
8.	Multiple Regression for Tardies for Language/Literature: All Grades	92
9.	Multiple Regression for Absences for Language/Literature: All Grades	93
10.	Multiple Regression for Tardies for Math: All Grades	94
11.	Multiple Regression for Absences for Math: All Grades	95
12.	Multiple Regression for Letter Grades for Language/Literature in Quarter 1: All Grades	96
13.	Multiple Regression for Letter Grades for Math in Quarter 1: All Grades	97
14.	Multiple Regression for Standardized Test Scores for Language/Literature: 11th Graders	98
15.	Multiple Regression for Standardized Test Scores for Math: 11th Graders	100

LIST OF FIGURES

Figure		Page
1.	Partial Theoretical Model of How SEL Leads to Better Academic Performance (Inside School Only)	102
2.	Theoretical Model of How SEL and Intelligence Lead to Better Academic Performance	103
3.	Logic Model for the Components of the Report Card	104
4.	Logic Model for the Components of the Report Card: Stengths of the Relationships	105

Introduction

The Need for a Major Shift in Mental Health Interventions

In order to assess the health-risk behaviors of youth, the United States (US) Centers for Disease Control and Prevention (CDC) has conducted the Youth Risk Behavior Survey (YRBS) with high school students biennially since 1991. According to the 2009 national report, 26.1% reported feeling so sad or hopeless nearly every day for at least two consecutive weeks that they stopped doing some usual activities during the past 12 months, 19.9% had been bullied on school property during the past 12 months, 31.5% had been in a physical fight in the last 12 months, and 24.2% had engaged in binge drinking (i.e., five or more drinks of alcohol within a couple of hours) during the past 30 days (Eaton et al., 2010). Thus, over one quarter of youth in the United States (US) have serious mental health problems.

Unfortunately, the mental health of adults in the US is not better. According to the World Health Organization's (WHO) World Mental Health (WMH) Survey Initiative, over 25% of the United States (US) population meets criteria for a DSM-IV mental disorder in any given year and almost 50% meet criteria in their lifetimes (Kessler et al., 2009). As the US has over 313 million people (http://www.census.gov/) and approximately 700,000 mental health professionals (Hoge et al., 2007, as cited in Kazdin & Blase, 2011), every active service provider would have to treat over 111 different people every year to reach everyone with a disorder.

After reviewing the above adult data, Kazdin and Blase (2011) argue that in order to reduce the burden of mental illness a major shift in intervention research and clinical practice must occur, particularly with our emphasis on individual psychotherapy as the

delivery model of choice. But, the conclusions apply equally to youth. One way to reduce the burden is to focus on prevention and promotion interventions as well as treatment interventions. Focusing on prevention and promotion interventions broadens the approach from mental illness (i.e., psychological disorders) to mental health well-being in general (Dalton, Elias, & Wandersman, 2007). Mental health well-being include social (e.g., sense of belonging), cognitive (e.g., self-efficacy), emotional (e.g., emotion regulation), and behavioral (e.g., bullying) aspects of life. While prevention and promotion work is important at all ages, such interventions are particularly important for children and adolescents. Not only do these interventions prevent and alleviate a lifetime of suffering; but according to the research by James J. Heckman, the Nobel Prize winner in Economic Sciences in 2000, investment in children makes sense even from a purely economic standpoint in terms of improved productivity (e.g., reduced medical costs and crime, increased number of skilled workers and wages) (Heckman & Masterov, 2007).

Schools are often a prime source for mental health work with youth as they spend much of their time there in large numbers. While schools may be a pragmatically wise choice in terms of intervention location, the traditional methods of addressing their mental health needs and the ratio of school mental health professionals to students is not pragmatically wise. Most interventions focus on direct intervention (e.g., crisis intervention, assessment, brief consultation, and referral). Meanwhile, the ratio of school psychologists and school social workers to students is 1 to 2,500 and the ratio for school social counselors to students is 1 to 1,000 (Adelman & Taylor, 1998, as cited in Kuperminc, Leadbeater, & Blatt, 2001). If these mental health workers only dealt with the students identified in the YRBS, each one would need to intervene with 250-625

students every year to help only the students with serious mental health needs. Thus, there is a need for a major shift in school-based interventions for mental health well-being, as well.

Using Social-Emotional Learning to Improve Mental Health and Academic Performance

Social-emotional learning (SEL) is one type of school-based prevention and promotion intervention (SBPPI) that can effective reach many students at once. While SEL can target at-risk students or individual classrooms, frequently it is applied on a universal level (i.e., school-wide). SEL refers to the "capacity to recognize and manage emotions, solve problems effectively, and establish positive relationships with others" (Zins & Elias, 2006, p. 1). The Collaborative for Academic, Social, and Emotional Learning (CASEL) (http://casel.org/), a international not-for-profit organization that promotes the science and evidence-based practice (EBP) of SEL, identifies five core SEL competencies: 1) self-awareness, 2) self-management, 3) social awareness, 4) relationship skills, and 5) responsible decision-making (see Appendix A for definitions and examples) (CASEL, 2005; Zins, Bloodworth, Weissberg, & Walberg, 2007). These are referred to as the "CASEL 5."

The theory of SEL is based on the understanding that many problematic behaviors are caused by similar risk factors (CASEL, 2005). Most modifiable risk factors and protective factors for psychopathology are commonly grouped under the categories of emotion dysregulation and interpersonal problems. These categories are intricate aspects of SEL, particularly with the CASEL 5 conceptualization. In support of this argument, a recent meta-analysis revealed that universal school-based interventions designed to

improve students' SEL led to increases in social emotional skills (g = 0.57; e.g., interpersonal problem solving, decision making), positive attitudes towards oneself and others (g = 0.23; e.g., self-efficacy, school bonding) and positive social behaviors (g = 0.24; e.g., getting along with others) as well as decreases in conduct problems (g = 0.22; e.g., disruptive class behavior, bullying) and emotional distress (g = 0.24; e.g., depression, anxiety, stress) (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

The effect size (ES) used in the above meta-analysis was Hedges' g (Hedges & Olkin, 1985), which is a more conservative ES than d. Traditionally, a value of .2 is small, .5 is medium and .8 is large (using Cohen [1992]'s conventions for d). While the effect sizes fell within the small range, with the exception of social emotional skills, it is important to interpret effect sizes in the context of prior research and practical value, as even small effect sizes can have large impacts when applied on a large scale (Durlak, 2009; Hill, Bloom, Black, & Lipsey, 2008; McCartney & Rosenthal, 2000). In that context, universal school-based SEL interventions are one way to create a major shift in the mental health well-being of youth.

Additionally, SEL theory recognizes that the best learning occurs through supportive relationships that make learning challenging and meaningful (CASEL, 2005). After reviewing the literature, CASEL (2005) reported that evidenced-based SEL programs have both direct and indirect impacts on student academic learning. By creating safe, well-managed, caring, and participatory learning environments and providing competence in the CASEL 5 skills, SEL interventions lead to greater attachment to schools, more positive development, and less risky behavior; all of which lead to better academic performance (e.g., graduation rates, academic grades. and test scores). In

support of this, the same meta-analysis referenced above also discovered that universal school-based SEL interventions lead to increases in academic performance (e.g., academic grades, standardized academic test scores) (Durlak, et al., 2011). The ES for SEL on academic performance ($g = 0.27$) found in the aforementioned meta-analysis is comparable to a recent meta-analysis of strictly educational interventions (Hill et al., 2007, as cited in Durlak, et al., 2011) and translates into an average of an 11% gain in academic achievement for both school grades and standardized tests across studies (Durlak, et al., 2011).

One possible specific pathway for SEL skills leading to academic achievement is illustrated through the empirical studies of Duckworth and colleagues (2005, 2011) which have demonstrated the importance of self-control for predicting academic success for middle school students from a wide variety of schools with various SES and ethnic compositions. Specifically, self-control is a larger predictor of grade point average (GPA) than IQ (Duckworth, et al., 2011; Duckworth & Seligman, 2005) and IQ is a larger predictor of standardized test scores than GPA (Duckworth, et al., 2011). As self-control is a critical component of the CASEL 5 category of self-management, self-management may be particularly important for the findings of the meta-analysis of Durlak et al. (2011).

An amalgam of the CASEL model (Figure 2 in CASEL, 2005) and the Duckworth and colleagues model (Figure 1 in Duckworth, et al., 2011) provides a comprehensive view for how SEL leads to better academic performance. First, a partial model will describe how SEL and experiences within the school lead to academic performance.

Second, a full model will describe how SEL, intelligence and experiences both within and outside school lead to academic performance.

In the partial theoretical model (see Figure 1), (a) SEL skills lead to (e) better academic grades through the combination of (b) greater attachment to school (e.g., school bonding, attendance), (c) better school behaviors (e.g., greater homework completion and prosocial behaviors, less disruptive class behavior and bullying), and (d) greater skills/knowledge acquired in school. The (b) school attachment and (c) school behaviors mutually reinforce one another; and the (c) better school behaviors also lead to (d) greater skills/knowledge acquired in school. Additionally, the (c) school behaviors, (d) skills/knowledge acquired in school, and (e) academic grades lead to (f) better standardized test scores.

Building upon the partial model, the full theoretical model (see Figure 2) adds two key elements. First, this model adds (h) cognitive intelligence as an explanatory factor for skills and knowledge acquired both (d) in school and (i) outside school. Second, it recognizes that (a) SEL skills also play a role in how a student responds to (g) events outside of school, which leads to either greater or fewer (i) skills and knowledge acquired outside school. For instance, a child who has greater attachment to and better behavior with people and organizations outside of school is more likely to experience more enriching activities (e.g., leadership positions, non-school educational opportunities). Of course, some of these (g) events experienced outside of school (e.g., poverty, violence) may negatively impact a child's (c) school behavior despite (a) strong SEL skills.

SEL skills in general and self-control in particular probably contribute more to academic grades than standardized test scores because learning that occurs in the school

environment is highly structured. Meanwhile, intelligence probably contributes more to standardized test scores because learning that occurs outside of the school is often less structured. Additionally, the creation process of standardized tests and intelligence is highly similar. For example, both rely upon rigorous psychometrics and pre-testing on large (e.g., statewide, national) samples.

The importance of the link between SEL and academic achievement cannot be overstated when it comes to the promotion of SEL to improve the mental health well-being of youth. In recognition of this, CASEL changed its name from the "Collaborative for the *Advancement* of Social and Emotional Learning" to the "Collaborative for *Academic*, Social, and Emotional Learning" after a meeting with school superintendents from around the US who told the CASEL Leadership Team that they were too focused on academic achievement to take time for SEL even when they wanted to do so (Elias & Moceri, 2012). Since then, CASEL has placed a greater emphasis on the academic link and its influence has grown.

The combination of the failure of the No Child Left Behind Act of 2001 (NCLB) with its strict focus on academic testing to achieve its desired results (Elias, 2009) and the growing link between SEL and academic achievement have motivated US governments to promote SEL. While eight states (i.e., Idaho, Illinois, Kansas, New York, Pennsylvania, Tennessee, Vermont, and Washington) have SEL standards at the K-12 level, only Illinois has free-standing comprehensive standards for SEL due to its 2004 legislation that required all school districts to develop SEL for their schools (http://casel.org/policy-advocacy/sel-in-your-state/). These SEL expansion efforts may soon be bolstered at the national level by the Academic, Social, and Emotional Learning

Act of 2011 (HR 2437) in the House of Representatives. This bipartisan bill would allow Title II funds (i.e., the Teacher and Principal Training and Recruiting Fund) to be used for teacher and principal professional development in SEL. HR 2437 defines SEL as the knowledge, attitudes, and skills that result from A) self-awareness and self-management, B) social awareness and interpersonal skills, and C) decision making skills and responsible behaviors. Notably, CASEL is a strong promoter of this bill and this bill's definition of SEL is highly similar to the CASEL 5, especially considering that CASEL often groups the two self skills together and the two social skills together. While it is uncertain if HR 2437 will become law, it does indicate a greater willingness at the federal level to view education in a broader context than standardized tests alone.

The Problem of Sustainability and Scalability for Assessment

If one accepts, in theory, the importance of SEL to mental health and academics, it stands to reason that it would be desirable to have an assessment of SEL in children that would be sustainable, scalable, and individualized. Assessment approaches that currently exist include a performance measure of emotional intelligence modeled after the Stanford-Binet intelligence test, a multi-page inventory of emotional intelligence in a Likert-scale format, 360 degree assessment systems, and numerous scales measuring a single component of SEL competence (Bar-On, 2007; CASEL, 2005; Elias, 2007; Hall & Hord, 2005; Horner et al., 2004; Wilson-Ahlstrom, Yohalem, DuBois, & Ji, 2011). While these approaches generally have good psychometric properties, they are not feasible for *ongoing* use for *all* children in educational settings. These approaches are costly in terms of time and money and have low sustainability and low scalability. Of course, these problems are not unique to SEL, but rather are a problem for all psychosocial

interventions. Nevertheless, the problems of assessment may be even more pronounced for schools that have to chronically deal with large budget cuts, overburdened teachers, and inconsistent data collection.

One solution put forth to address the assessment problem is to make our measurements at the school level rather than the individual level. For instance, instead of measuring individual level SEL skills by self- or teacher-reports, we could measure school level SEL skills by measuring the school climate. School climate consists of the quality of interpersonal interactions within schools, which includes such concepts as how much students show respect for one another, how supportive and caring teachers are, and how involved students are in shaping their school environment (Lickona & Davidson, 2003; Mattison & Aber, 2007). Schools that have higher levels of these behaviors have higher levels of individual SEL. The Developing Safe and Civil Schools (DSACS) Project was one such intervention. DSACS aimed at improving school climate and academics and decreasing violence through implementing SEL in schools. DSACS used anonymous climate surveys to assess its impact on over 10,000 students from 250 schools (Elias, 2009; Moceri, Elias, Fishman, Pandina, & Reyes-Portillo, 2012). Unfortunately, this approach lacks the ability to link individual SEL to individual academic achievement, and it can only identify group level trends in SEL rather than the trends of specific individuals.

Report Card Comments as SEL Assessment

While there is currently no recognized feasible, scalable assessment system for SEL, there is an assessment system in educational practice already utilized in the vast majority of schools: the report card. If one looks at report cards, one often will find on the

other side of the academic grades a set of comments about children's behavior, character, preparation, motivation, and more. In other words, in ways that seem far from systematic, elements of SEL already are being assessed on report cards. Interesting, SEL is often referred to in the SEL literature as the skills that reflect "the other side of the report card" (Elias, Wang, Weissberg, Zins, & Walberg, 2002), even though no study to date has systematically and empirically studied how the behavior ratings and comments sections of the report card relate to SEL, academic grades, or standardized tests. This study seeks to be the first one to do so, with a particular interest on those comments that reflect the presence or absence of SEL skills. If these comments are indicative of current and future academic success, then a new methodology for assessing individual students in a sustainable and scalable can be created. That is, the comments on the report cards can become a tool for assessing every student's SEL skills.

One reason that report card comments were chosen is that the vast majority of teachers already assign comments (i.e., scalability is achieved) to each student (i.e., individualization is achieved), which means that this type of tracking could provide useful guiding information about future life and academic success (i.e., pragmatic utility is achieved) without putting an additional burden on teachers (i.e., sustainability is achieved). This is particularly important as schools rarely collect psychological measurements when not currently engaging in research efforts with a university or non-profit organization. Another reason that report card comments were chosen is that while millions of dollars are spent on the evaluation of students' academic abilities and achievements, little work has been done evaluating the validity of report card comments even though a substantial amount of the conversations between parents and teachers

focus on students' behaviors. A third reason is that program evaluators and researchers could one day use report card comments to longitudinally evaluate students on a quarterly basis before, during and after an intervention.

If, however, report card comments are going to be used as an SEL assessment, then a logic model for the various components of the report card needs to be derived. Therefore, a logic model (see Figure 3) is presented below based upon the theoretical models outlined earlier (see Figure 1 and Figure 2). In this logic model, (a) previous comments lead to (e) current academic grades through a combination of (b) attendance, (c) current comments, and (d) previous academic grades. The (c) current comments also lead to (b) better attendance. Finally, (f) better standardized test scores are partially achieved through (c) current comments, (d) previous academic grades, and (e) current academic grades. Similar to the theoretical model, comments are tied more to academic grades than standardized achievement test scores due to the same reasons previously given. The same lettering was used for the logic model as the theoretical models to aid in comparisons (see Appendix B for details).

Other report card studies. Despite an exhaustive literature search, only two other research teams were discovered that have empirically studied report card comments. The first team, headed by Stephen J. Friedman, examined the characteristics of report cards (Friedman & Frisbie, 1995) and teachers' use of computerized report card comments (Friedman, Valde, & Obermeyer, 1998) using Wisconsin report cards. While most elementary school report cards had space for teacher-written comments, almost all of the high school report cards relied exclusively upon computer menus for comments. Middle school report cards used both methods about evenly. The number of

computerized comments typically ranged from 50 to 100, and teachers were often restricted to a maximum of two comments per student per subject area for each reporting period (Friedman & Frisbie, 1995).

Friedman and colleagues' second study (1998) focused exclusively on computerized report card comments and examined a small town middle school of 475 students in southeastern Wisconsin. This school allowed teachers to select two comments from a menu of 82 total comments. These comments were listed under five general headings: positive academic, positive behavioral, negative academic, negative behavioral, and general. Only 17% of academic grades did not have an accompanying comment; but this varied substantially across teachers (with some teachers assigning no comments) and only 52% of students received two comments. A strong majority (72%) of teachers viewed two comments as "about right," but less than half (42%) of parents agreed with them; the remainder of respondents believed that two comments per course were "not enough". The teachers (n = 37) used an average of 17 different comments, which is only 20.7% of their options, indicating that 82 options may be impractical. While the teachers were more likely to assign positive comments than negative comments in general, teachers were more likely to use negative comments to explain low grades than positive comments to explain high grades.

The studies by Friedman and colleagues (1995; 1998) are different from the this study because their work was an atheoretical examination of how report cards are used. Meanwhile, this study seeks to examine the relationship between report card comments and academic achievement from a SEL lens. Friedman and colleagues have not

conducted any further research on this topic and are unaware of similar work (G. A. Valde, personal communication, February 10, 2012).

The second team developed New York City's first-ever character report card (Tough, 2011). This team is headed by Angela Lee Duckworth and Christopher Peterson. They designed the report card working with the local superintendent and headmaster, which is important for creating buy-in and sustainability. The character report card is distributed to students at the KIPP network of charter schools twice a year and consists of 24 indicators to measure seven strengths identified through the positive psychology literature: zest, grit, self-control, social intelligence, gratitude, optimism and curiosity. This report card could be used to create a "character-point average" (CPA) to compliment the more traditional grade-point average (GPA). While the KIPP report card is being studied empirically, no studies have been published yet on this data (A. L. Duckworth, personal communication, November 16, 2011). The self-control studies by Duckworth and colleagues cited earlier in this dissertation relied upon self, parent, and teacher survey reports.

The proposed research here is similar to the research around the KIPP character report card in that both examine report card comments, but it differs in two key ways. First, this study will look at the comments through the SEL lens of the CASEL 5 (i.e., self -awareness, self-management, social awareness, relationship skills, and responsible decision-making) instead of through a positive psychology lens. Second, the present study uses a bottom-up approach of studying report card comments that already exist rather than using a top-down approach of creating report card comments and then studying them.

Case study approach. This study is part of a larger study, the Social-Emotional Learning Indicators Project (SEL-IP), in which this author is co-principal investigator. The purposes of the project reflect the guiding framework of this dissertation. Because the SEL-IP utilizes a bottom-up approach, this project lends itself to a blend of the positivist and pragmatic paradigms. The positivist paradigm contends that physical and social reality can be quantified through objective, scientific study, which is gained through theory-embedded, hypothesis testing experiments, which emphasize context-free, universal laws (Fishman, 1999; Fishman, Rego, & Muller, 2010). Meanwhile, the pragmatic paradigm emphasizes case-based knowledge that is contextually embedded and usually employs both quantitative and qualitative data (Fishman, et al., 2010). Within the pragmatic paradigm, the case can be an individual (e.g., a student) or an organization (e.g., a school or district).

By using quantitative measurements of academic achievement (i.e., letter grades and standardized test scores), by using report card comments that are chosen by teachers from predetermined lists (by the school and/or district), and by controlling for demographic variables, the SEL-IP embraces the positivist paradigm. However, by explicitly acknowledging that the academic achievement and comments are socially constructed, by examining the context of the schools and districts analyzed, and by viewing each school and/or district as an individual case that contains unique contextual information, the SEL-IP embraces a pragmatic paradigm. Additionally, the very fact that the SEL-IP utilizes report cards as part of its design is very practical. A primary goal of the pragmatic paradigm is to solve context specific problems, so a study cannot be pragmatic if it is not also practical. While studies using the positivist paradigm can be

practical as well, it is not a requirement as its primary mission is to derive theory-based general laws.

The SEL-IP as a whole is best conceptualized as a series of case studies, where each case is a specific school and/or an entire school district. The approach is to look at a variety of report cards in order to determine which types of comments have most pragmatic utility for understanding academic success and promoting SEL in students of all ages. For instance, one SEL-IP case study, using data from 600 middle school students from a district in New Jersey, found that minority (i.e., Black and Hispanic) students and male students received more negative comments, less positive comments, and worse academic grades than their White and female peers (Kemp, Moceri, & Elias, 2012). As that case study used three middle schools, there is some support for generalizability; but it is currently unknown if other districts with different report card comments would follow a similar pattern. If this dissertation (with its use of a separate school district, a completely different set of report card comments, and a different grade level [i.e., high school]) can replicate the Kemp et al. (2012) results, our confidence in the generalizability of those findings increases. Alternatively, if the results are not replicated, contextual differences can be examined to see why dissimilar results emerged (e.g., differences in the ethnic composition of the school or specific comments utilized). Thus, while no single case study of the SEL-IP will be definitive, after a series of case studies have been conducted, synthesis of knowledge can be ascertained by examining commonalities and differences.

Research questions (RQ) and hypotheses (H).

This study will examine three research questions in detail: 1) What conceptual categories can be created using report card comments?, 2) What does the distribution of report card comments look like?, and 3) How well do the report card comments adhere to SEL theory? Below are the details of these questions and the specific hypotheses that pertain to them.

RQ1: What conceptual categories can be created using report card comments? Are the CASEL 5 dimensions of SEL represented well by the report card comments?

RQ2: What is the distribution of report card comments for different levels of academic performance, different demographic groups, and different grade levels (e.g., 9th-12th grade)? Can students with low academic grades receive positive comments? Can students with high academic grades receive negative comments?

H2a: Students with an at-risk demographic profile will have worse outcomes (e.g., Kemp, et al., 2012). An at-risk demographic profile is defined as being an ethnic minority, being male, or having a low SES. Importantly, this dissertation and the cited research predicate ecological biases in US sociopolitical history as the cause of these demographic differences (see Elias & Moceri, 2012 for details). Worse outcomes are defined as receiving more negative comments, less positive comments, lower academic grades, and lower standardized test scores than their non-at-risk peers.

H2b: Students with low academic grades are more likely to receive comments (e.g., Friedman, et al., 1998).

RQ3: How well do comments on student report cards adhere to SEL theory (see Figure 2 and Figure 3)?

H3a: More SEL skills (represented by report card comments) will be associated with fewer attendance problems (represented by tardies and absences) (e.g., Duckworth & Seligman, 2005).

H3b: More SEL skills will be associated with better concurrent academic performance (represented by letter grades and standardized tests) (e.g., Duckworth & Seligman, 2005; Durlak, et al., 2011).

H3c: More SEL skills will predict future academic success (represented by letter grades and standardized test scores), even after controlling for prior academic success (represented by letter grades) (Duckworth, et al., 2011).

The hypotheses listed above also are expected to hold true for both SEL-type and non-SEL-type comments on the report cards, but SEL-type comment findings are expected to be stronger. As the pathways linking SEL to academics is believed to function through common outcomes (see Figure 2) rather than through the promotion of enhancements of subjective skills, the hypotheses are expected to be true for both classes typically viewed as having more "subjective" grading criteria (e.g., language/literature, social studies) and those with more "objective" criteria (e.g., math, science). Finally, the hypotheses are expected to be stronger for classroom academic success (i.e., letter grades) than for standardized academic success (i.e., standardized test scores) (Duckworth, et al., 2011).

Method

Setting and Participants

Participants are students from a high school in New Jersey with an enrollment of approximate 2,000 students from the 2007-2008 academic school year. Approximately

20% of students qualify for a free or reduced lunch, even though the school is located in a high socioeconomic status (SES) district. Over 50% of the students are Black and over 33% are White.

The school used for this dissertation is appropriate for an SEL-IP case study because it (a) has participated in Rutgers' Developing Safe and Civil Schools (DSACS) Project, which supports locally-developed approaches to improving school climate and promoting students' social-emotional and character development, (b) uses an infused approach to SEL rather than a specific program approach, which means that the school's efforts are more likely to reflect the CASEL 5 as a whole rather than only those aspects related to a unique program, (c) has a superintendent with SEL expertise and leadership experience, (d) has no formal or consultative relationship with CASEL, and (e) uses computerized report card comments in a manner similar to the findings of Friedman et al. (1995, 1998). Finally, the use of a high school is deliberate as there is not as much SEL research at the high school level (e.g., only 13% of the studies in the Durlak, et al., 2011 meta-analysis). The DSACS Project was conducted after appropriate review from the Institutional Review Board for the Protection of Human Subjects (IRB) of Rutgers, The State University of New Jersey.

Inclusion criteria for this study's analyses were all White and Black students who took a year-long language/literature course (LA) and/or a year-long math course (MA) during the 2007-2008 academic school year. Exclusion criteria were students of other ethnicities (about 10%; as combining them into one group could be highly misleading) and students who have an individualized education program (IEP) (about 10%; as their grades and the relationship between their grades and their comments may be substantially

different from mainstreamed students). Language/literature and math courses were chosen because these subject areas have standardized test scores in the state of New Jersey and most high school students take these classes. These criteria resulted in 1,247 unique students. Of the 239 students who qualified for a free or reduced lunch, only 4 (1.7%) were White. As one cannot generalize from this tiny percentage, these 4 students were eliminated from the dataset.

Thus, this dissertation examined 1,243 unique students. Of these, 52.5% were female, 57.8% were Black, and 18.9% qualified for a free or reduced lunch (which was 32.7% of the Black sample). Students were evenly split across the grade levels, χ^2 (3, N = 1243) = 3.99, p = .262). The specific demographics and sample sizes for the four main types of analyses (i.e., language/literature classes, math classes, language/literature standardized test scores, and math standardized test scores) are presented in Table 1. When a student had more than one year-long class in language/literature (n = 36; 3.4% of those with at least one year-long language/literature class) or math (n = 46; 5.4% of those with at least one year-long math class), one of the classes was randomly selected for analysis using Random.org, a true random number generator (Haahr, 2006).

Measures

Report card variables. Academic letter grades and behavior comments are provided for each quarter for each class as part of the regular report cards for students at this school. Letter grades are provided separately for the academic subject areas of language/literature and math. Letter grades range from A+ to F, where A+ or A = 4.0, A- = 3.67, B+ = 3.33, B = 3.0, B- = 2.67, C+ = 2.33, C = 2.0, C- = 1.67, D+ = 1.33, D = 1.0, D- = 0.67, F = 0.

Each student can receive up to two behavior comments per class per quarter. Teachers use these comments to provide feedback about a wide range of areas, including social behavior, motivation for learning, study skills, and character. They chose from a list of 25 possible behavior comments (see Table 2). Ten comments were from of one list and fifteen comments were from another list.

Key school informant interview. In order to better understand the use of report card comments an interview was conducted with a key school informant, who had been at the school over 20 years, was a teacher for over 10 years, and at the time of the interview was currently the chair of a subject area. The informant provided the information in the below paragraph based upon the informant's years of experience (personal communication, March 16, 2012).

The comments were created and modified by various committees of teachers over the years. Comments are selected based upon the closest fit to the student and the comments can be viewed by other teachers. Different subject areas use the comments in the same manner. Most teachers take the comments seriously and view them as conversation starters with parents and students.

Non-report card variables. Students' grade level, SES, IEP status, absences, tardies, and standardized test scores were provided from students' regular school records. Students' SES was measured by eligibility for free or reduced lunch (0 = no, 1 = yes), which is determined by federal guidelines. Grade level was dummy coded with 11th grade being the reference group as only 11th graders had standardized test scores.

The High School Proficiency Assessment (HSPA) is New Jersey's standardized achievement test for high school students. Scaled scores range from 100-300, where 100-

199 means Partially Proficient, 200-249 means Proficient, and 250-300 means Advanced Proficient. Students take the HSPA in March of their junior year; therefore, standardized test score analyses are limited to this grade level.

Effect Sizes

An *r* of .1 is small, .3 is medium and .5 is large for correlations (Cohen, 1992). A change in R square (ΔR^2) value of .02 is small, .13 is medium and .26 is large for multiple regression; the same interpretations are used for partial eta squared (η_p^2) values for analysis of variance (ANOVA) (Cohen, Cohen, West, & Aiken, 2002). A Hedges' *g* of 0.2 is small, 0.5 is medium, and 0.8 is large, using Cohen's (1992) conventions for *d* for independent-samples *t* tests. Hedges' *g* was calculated using DeFife (2009). All other statistics were calculated using SPSS Statistics v.19.

Results

On average, students were assigned grades in the C+ to B- range, missed about three weeks of classes, were tardy a week, scored in the proficient range on their standardized tests, and received one comment per quarter. Table 1 provides the specific averages, standard deviations, counts, and percentages for these variables. For brevity, academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively.

There was a large correlation between LA and MA letter grades when looking at all grade levels for Q1, $r(658) = .51$, $p < .000$, which is a large ES. The r value was .51, .52, and .51 for Q2, Q3, and Q4 respectively. There was also a large correlation between LA and MA standardized test scores, $r(135) = .64$, $p < .000$, which is a large ES.

Repeated measures ANOVAs revealed a decrease in letter grades and the number of comments over time. LA letter grades in Q4 were lower than Q1 to Q3. MA letter grades in Q2 to Q4 were lower than Q1 (see Appendix C for details). Similarly, fewer comments were given in Q4 for both LA and MA (see Appendix C for details). Across the quarters for LA, the weakest correlation was between Q1 and Q4, $r(1052) = .68$, $p < .000$ and the strongest correlation was between Q1 and Q2, $r(658) = .76$, $p < .000$; both were large ES. Across the quarters for MA, the weakest correlation was between Q1 and Q4, $r(847) = .68$, $p < .000$ and the strongest correlation was between Q1 and Q2, $r(847) = .76$, $p < .000$; both were large ES.

RQ1: Conceptual Categories for Report Card Comments?

Grounded Theory (Corbin & Strauss, 2008) was used to see how well report card comments matched the CASEL 5 conceptualization of SEL skills. All raters were given written instructions to independently determine whether each of the report card comments were representative of the CASEL 5 domains or not. As many report card comments reflect problematic behaviors, comments were assigned to a CASEL 5 domain whether they represented the presence or absence of a specific SEL skill. Coders were provided with a document with definitions and examples of the CASEL 5 domains (see Appendix A) as well as the list of 25 report card comments (see Table 2).

Category Creation. A three-phase process was used to create report card comment variables. In phase one, two SEL-experienced undergraduate research assistants (RAs) provided ratings and reached a consensus. The RAs assigned each comment to one of six categories: Either to one of the CASEL 5 domains (i.e., self-awareness, self-management, social awareness, relationship skills, and responsible decision making) or to

a non-SEL category. While 10 comments were assigned to a CASEL 5 domain, 5 of them were assigned to self-management and the other 5 were spread across the other four domains. Therefore, the four non-self-management CASEL 5 domains were combined into one category: Other SEL.

In phase two, two school psychologists who are SEL experts provided ratings and reached a consensus. The school psychologists were given the same instructions as the RAs, except that they were told to assign each comment to one of three categories: Self-management comments (a CASEL 5 dimension of SEL), other SEL comments (CASEL 5 dimensions of SEL besides self-management), and non-SEL comments. The consensus between the two RAs, who were counted together as one source, and the two school psychologists, who were counted as one source each, is presented in Table 3.

In phase three, this author separated the non-SEL category into four subcategories and then divided the six categories (i.e., two SEL ones and four non-SEL ones) into positive comments (e.g., presence of skills) and negative comments (e.g., absence of a skills) based upon the wording of the comments. This resulted in 10 comment categories that were created a priori for analyses in this dissertation (see Table 4).

CASEL 5 representation. Ten of the 25 report card comments were included in one of the two SEL comment categories. Five of the 10 SEL comments represented the presence or absence of the core SEL competency of self-management, so they were given their own comment category. The other five SEL comments were put into another category (i.e., other SEL) to represent the presence or absence of the other four core competencies of the CASEL 5. However, it should be noted that while there was enough of a representation of the CASEL 5 in report card comments to proceed with subsequent

analyses, the mapping was clearly partial and did not allow for a specific analysis of how each of the CASEL 5 dimensions link to academics.

Frequency of categories. Although many students received two comments in the same quarter (i.e., 11.3% to 42.5% for LA classes; 5.6% to 27.9% for MA classes), very few students received two comments within the same comment category (i.e., 0% to 4.7% for LA classes; usually 0% to 1.5% for MA classes). Due to the rarity of these "double counts," all comment categories were recoded to have a maximum value of one in order to make statistical inferences easier for all analyses. Thus, specific comment categories have a can have a minimum of zero and a maximum of one.

Across the quarters and analyses, four comment categories often had frequency percentages in the tens and twenties: positive self-management (minimum [min] = 3.6% and maximum (max) = 24.0%), negative self-management (min = 4.1%; max = 26.2%), positive other SEL (min = 4.8%; max = 25.6%), and positive preparedness/interest (min = 1.5%; max = 32.9%). Four comment categories often had frequency percentages in the middle and high single digits: positive academic/extremes/other (min = 1.5%; max = 8.2%), negative academic/extremes/other (min = 1.5%; max = 15.0%), negative preparedness/interest (min = 3.1%; max = 9.4%), and improvement/ambivalence (min = 0.9%; max = 11.3%). Finally, four comment categories often had frequency percentages in the low single digits: negative attendance (min = 2.4%; max = 6.2%) and negative other SEL (min = 0.0%; max = 2.5%). The frequency percentages for LA for all grades are presented in Table 5 for each of the 10 categories for each of the four quarters. The results were similar for LA for 11th graders, for MA for all grades, and for MA for 11th graders.

Notably, three of the four most common categories are SEL categories and the other one was labeled as SEL by two of the three rating sources. Additionally, the frequency counts were not a mirror image of how many comments were in each category. For instance, positive self-management only has only one comment but it was one of the most used categories. Meanwhile, negative attendance has three comments but it was one of the least used categories.

Relationship between categories. In order to better understand the relationship between the 10 different comments, the correlation between the comments categories were examined when students had two comments in the same quarter. As all values are either zero or one, there are many tied values (i.e., ranks). Kendall's tau-b was used as it is a non-parametric correlation that is equipped to deal with a large number of tied ranks. A positive correlation can be interpreted to mean that the combination of these two comment categories is likely (i.e., when one is present the other is also present). A negative correlation can be interpreted to mean that the combination of these two is unlikely (i.e., when one is present the other is not present).

First, though, agreement between the two academic subject areas (i.e., LA and MA) was examined for the comments categories. While there was very strong agreement between LA and MA for letter grades and standardized test scores, the agreement was not as strong for the comments. For example, the correlation between LA and MA for positive self-management in Q1 for all grade levels, $\tau(658) = .18$, $p < .000$, was weaker than the correlation for their letter grades; and the correlation between LA and MA for negative self-management in Q1 for all grade levels, $\tau(658) = .24$, $p < .000$, was weaker than the correlation for their letter grades. Many other comment category correlations

were even lower (e.g., positive preparedness) or were not significant (e.g., positive other SEL). The weaker correlations for the comments was true even after the letter grade correlations were reanalyzed with Kendal's tau (τ) instead of Person's *r*. Therefore, subsequent questions were run separately for each subject area.

As the report card structure only allowed two of the ten categories to be anything but zero and as many comment categories had low frequencies, most correlations were either negative or non-significant. When examining all grade levels together, positive other SEL and positive preparedness/interest were positively correlated for all four quarters for both LA and MA ($\tau_{min} = .22$; $\tau_{max} = .54$) and negative self-management and negative preparedness/interest were positively correlated for the first three quarters for LA and the first quarter for MA (when significant: $\tau_{min} = .32$; $\tau_{max} = .54$). Notably, two of the three rating sources for creating the comment categories viewed preparedness/interest as an SEL category. Thus, even though the preparedness categories are not a part of the CASEL 5, the CASEL 5 SEL skills appeared to be associated with them both conceptually and empirically. Meanwhile, all other comment categories were relatively independent from each other.

RQ2: Distribution of Comments by Academic Performance?

Several analyses were performed in order to better understand how the comments were distributed in general. First, the general comments were analyzed by grade level (i.e., 9th through 12th). Second, t-tests were run to determine the average letter grades with positive, negative and ambiguous comments. Third, the frequencies chi-squares of the general comments of positive, negative, and ambiguous were examined by letter grade (i.e., A through F).

The positive comments include all of 4 positive categories (i.e., positive self-management, other SEL, academic/extremes/other, and preparedness/interest). The negative comments include all 5 negative categories (i.e., negative self-management, other SEL, academic/extremes/other, attendance, and preparedness/interest). The ambiguous comments category includes only the improvement comment category from the final 10 categories created.

Grade level. Repeated measures ANOVAs revealed that grade level sometimes had a small ES on letter grades and comments. Specifically, students in 9th grade were more likely to receive lower MA letter grades, more positive comments in LA, more negative comments in MA, and more ambiguous comments in LA and MA (see Appendix D for details). Additionally, 12th graders were less likely to receive positive comments in LA. Other overall patterns were not observed. Therefore, grade level will be entered into all multiple regression equations to control for these effects.

H2a: At-Risk Profile.

Repeated measures ANOVAs revealed that students with an at-risk profile, as defined as being a Black student, being male, or qualifying for a free or reduced lunch, had worse outcomes. Black students had lower letter grades in LA and MA, lower HSPA scores in LA and MA, fewer positive comments in LA and MA, more negative comments in LA and MA, and more ambiguous comments in LA and MA; the effect sizes for letter grades were medium, for HSPA scores were large, and for the three comments types were small (see Appendix E for details). Male students had lower letter grades in LA and MA, lower HSPA scores in LA, fewer positive comments in LA, more negative comments in LA and MA, and more ambiguous comments in LA; all effect sizes were small (see

Appendix F for details). Students qualifying for a free or reduced lunch had lower letter grades in LA and lower HSPA scores in LA and MA; the ES for LA letter grades was less than small, for LA HSPA scores was medium, and for MA HSPA scores was small (see Appendix G for details). Therefore, ethnicity, gender, and lunch status were entered into all multiple regression equations to control for these effects.

H2b: Lower Grades when More Comments.

Frequencies. The frequencies and chi-squares of the general comments were examined to determine if students with positive comments could receive low letter grades, to determine if students with negative comments could receive high letter grades, and to determine what letter grades were most common for the ambiguous comments. To examine this, the letter grades were recoded into the five key variables (i.e., As, Bs, Cs Ds and F without the pluses or minuses) and the comments were recoded to have a maximum value of one. All chi-square analyses were significant. For all LA analyses χ^2 (4, N = 1054) = 22.03 to 487.64, $p < .000$, except for Q4 for ambiguous comments when the results were χ^2 (4, N = 1054) = 9.86, $p = .043$. For all MA analyses χ^2 (4, N = 849) = 33.76 to 357.41, $p < .000$. This indicates that the general comments were not equally distributed among the letter grades for any quarter for LA and MA.

The minimum and maximum percentages for each general comment for each letter grade across the four quarters are presented in Table 6. Overall, positive comments were mostly associated with As and Bs but Cs were also frequent. Negative comments were mostly associated with Fs, Ds, and Cs but Bs were also frequent. Ambiguous comments were mostly associated with Ds and Cs but Fs and Bs were also frequent. These findings occurred for both LA and MA classes. Nevertheless, every general

comment category occurred for each of the five key letter grades for at least 1% of the cases in at least one quarter. Therefore, even though that was the most common pattern was for positive comments to occur with high letter grades and negative comments to occur with low letter grades, this pattern was not deterministic.

Positive Comments. Overall LA positive comments are presented in Table 7. Violations of Levene's test of equality of variance required independent t-tests for each comparison, using .0167 (.05/3) for the alpha for these analyses. For LA analyses in Q1, those having two positive comments ($M = 3.47$, $SD = 0.47$) received higher letter grades than those having one comment ($M = 3.04$, $SD = 0.81$), $t(522.64) = 7.72$, $p < .000$, $g = 0.58$, which is a medium ES, and higher letter grades than those having zero comments ($M = 1.91$, $SD = 1.19$), $t(650.72) = 24.09$, $p < .000$, $g = 1.48$, which is a large ES. The difference between having one comment and having zero comments was also significant, $t(856.10) = 16.85$, $p < .000$, $g = 1.10$, which is a large ES. This means that for Q1 those with two, one and zero positive comments averaged in the B+, B, and C ranges, respectively. Similar findings for LA were found in Q2, Q3, and Q4 as in Q1.

The overall number of MA positive comments was reasonably similar to the means, standard deviations and percents presented in Table 7 for LA positive comments, with the exception that very few students (1.1%-4.1%) received two comments in Q2, Q3, and Q4. Therefore, the Q2 through Q4 results were only analyzed for having one or more comments versus having zero comments. The MA Q1 results were similar to the LA Q1 results. For MA analyses in Q2, those having one positive comment ($M = 2.99$, $SD = 0.78$, $n = 301$) received higher letter grades than those having zero comments ($M = 1.97$, $SD = 1.14$, $n = 533$), $t(803.54) = 15.28$, $g = 0.99$, which is a large ES. Thus, those

with one positive comment averaged in the B range whereas those with zero comments averaged in the C range. Similar findings for MA were found in Q3 and in Q4 as in Q2.

Negative Comments. Overall LA negative comments are presented in Table 7. Violations of Levene's test of equality of variance required independent t-tests for each comparison, using .0167 (.05/3) for the alpha for these analyses. For LA analyses in Q1, those having two negative comments (M = 1.01, SD = 0.93) received lower letter grades than those having one comment (M = 1.60, SD = 1.00), $t(313) = -5.19$, $p < .000$, $g = -0.60$, which is a medium ES, and lower letter grades than those having zero comments (M = 3.10, SD = 0.78), $t(145.58) = -23.17$, $p < .000$, $g = -2.61$, which is a large ES. The difference between having one comment and having zero comments was also significant, $t(260.99) = -19.56$, $p < .000$, $g = -0.67$, which is a medium ES. This means that for Q1 those with two, one and zero negative comments averaged in the D, C-, and B ranges, respectively. Similar findings for MA were found in Q2, Q3, and Q4 as in Q1, except in Q4 the grades were one-third of a letter grade lower (i.e., D-, D+, and B- instead of D, C- and B) for each group.

The overall number of MA negative comments was reasonably similar to the standard deviations and percents presented in Table 7 for LA positive comments. The means for Q1 through Q4 were 0.29, 0.35, 0.32 and 0.19, respectively. The MA Q1 results were similar to the LA Q1 results. The MA Q2 and MA Q3 results were similar to the LA Q4 results. In MA Q4 very few students had two negative comments (n = 21, which is 2.5%), so only those with one or zero comments were examined. For MA Q4 those having one comment (M = 2.59, SD = 1.08, n = 123) received lower letter grades than those with zero comments (M = 1.24, SD = 1.08, n = 705), $t(826) = -12.71$, $p < .000$,

g = -1.24, which is a large ES. Thus, those with one negative comment in MA Q4 averaged in the C+ range and those with zero negative comments averaged in the D range.

Ambiguous Comments. Overall ambiguous comments for LA are presented in Table 7. For LA Q1 those with at least one ambiguous comment (M = 1.61, SD = 0.79, n = 51) received lower letter grades than those with zero comments (M = 2.64, SD = 1.15, n = 1003), $t(61.34) = -8.75$, $p < .000$, g = -0.90, which is a large ES. This means that for Q1, those with an ambiguous comment averaged in the C- range whereas those without one averaged in the B- range. Similar findings for LA were found in Q2, Q3, and Q4 as in Q1, except that the averages were closer to C and C+.

Overall ambiguous comments for MA were about double the rate of those for LA. Specifically, the percentage of students with at least one ambiguous comment for MA were 8.5%, 9.1%, 11.3%, and 5.3% in Q1 through Q4, respectively. The MA results for all four quarters were similar to the LA Q1 results, except that the MA Q4 letter grades were in the D+ and C+ ranges.

RQ3: Comments and SEL Theory?

The three hypotheses for this question were tested using multiple regression analyses. These were run using the forced entry method for the steps. This method is better to use when theory testing as the stepwise method can be unduly influenced by random variation in the data, which decreases the possibility of replication (Field, 2009). After the initial analysis was run, steps that were not significant ($p > .05$ for ΔR^2) were eliminated from the model. Likewise, when coefficients were not significant in their

initial step, they were eliminated from the model. This process of removing statistical redundancies is recommended by Field (2009) and frees up degrees of freedom.

Two exceptions were made to the above process. First, all grade level variables were kept, as long as the step was significant, even if one or two coefficients were not significant, because each was part of the same dummy coding set. Second, when the attendance count variables (i.e., tardies and attendances) were the dependent variable (DV), the attendance comments were always kept in the model, even if they were not significant, because they explicitly referenced the DV.

H3a: SEL Skills Are Associated with Fewer Attendance Problems.

Analyses of the attendance count variables (i.e., tardies and absences) revealed strong skewness and kurtosis (e.g., skewness = 3.15 and kurtosis = 12.42 for tardies and skewness = 3.62 and kurtosis = 23.44 for absences for LA). Therefore, values greater than the 95th percentile were recoded to one unit higher than the 95th percentile after centering them. This dramatically reduced skewness and kurtosis (e.g., skewness = 1.87 and kurtosis = 2.62 for tardies and skewness = 0.61 and kurtosis = -0.35 for absences for LA). The recoding of extreme values is a conservative approach that minimizes their effects on parametric statistics while maintaining their interpretability (Tabachnick & Fidell, 2006).

Multiple regression analyses were performed with tardies and absences being the dependent variables (DVs), where step 1 was the demographic variables (i.e., gender, ethnicity, SES, and their interactions), step 2 was the grade level, step 3 was the comment categories specifically related to attendance for all quarters, and step 4 was the remaining comment categories for all quarters.

This method was chosen as attendance was only available for the total year (i.e., not by quarters) and this data analysis plan produced the most parsimonious and conservative interpretation. Exploratory analyses with LA tardiness revealed that analyzing each quarter as a separate step in one regression produced virtually identical results and that analyzing the comments for each quarter in separate regressions produced similar results that were slightly more liberal (i.e., more comment categories were significant).

The ΔR^2, *b*, *SE b*, and β values for the four attendance analyses are presented in Tables 8 through 11. As all independent variables were on the same dichotomous metric, the *b*-values can be compared directly. A positive *b* indicates how many additional tardies or absences a student had above the average while a negative *b*-value indicates fewer tardies or absences. The results in general showed that demographics had a small ES on attendance, grade level had a small ES, attendance comments had a medium ES, and non-attendance comments had a small ES. White students had fewer tardies in both subjects but more absences in both subjects. Similarly, female students had more tardies in MA but more absences in both subjects. Ninth and 10th graders had fewer tardies and absences than 11th graders.

Often non-attendance comments in all four quarters were significant. The negative comments were associated much more often with attendance than positive comments. The most frequently significant non-attendance comment was negative self-management; as its *b*-value ranged from 1.73 to 2.58, students who had this comment were tardy or absent approximately 2 additional times over the course of the academic year after controlling for all other variables. Two comments, negative other SEL and

negative academics, had both positive and negative *b*-values, which makes their results difficult to interpret. Notably, students who had the ambiguous improvement comments had a positive *b*-values, indicating worse outcomes.

In order to determine how the 95th percentile data transformation influenced the above results, two alternative data transformations for the attendance count variables were tested: taking the square root and taking the log plus one. Both of these bring larger scores closer to the center and reduce positive skew (Field, 2009). Those results were highly consistent with the results presented earlier, with the exception that their *b*-values were not interpretable.

H3b: SEL Skills are Associated with Better Academics.

Multiple regression analyses were performed with the letter grades being the dependent variables (DVs), where step 1 was the demographic variables (i.e., gender, ethnicity, SES, and their interactions), step 2 was the grade level, step 3 was the attendance variables (i.e., tardies, absences, and attendance comments), and step 4 was the remaining comment categories. The letter grades were "centered" to a B- value (2.67), which approximates the mean value. This time all quarters were analyzed separately.

Attendance comments were put in the same step as the attendance count variables (i.e., tardies and absences) as they measure the same construct. Exploratory analyses revealed that excluding the attendance count variables (which are year-long rather than quarter variables) did not have a meaningful effect on the results, so they were kept in the analyses. All variables, with the exception of the attendance count variables, are dichotomous, so again the *b*-values were interpreted instead of the βs. A *b*-value of 0.33 indicates an improvement of a third of a letter grade above the average (e.g., going from a

B- to a B), a *b*-value of 0.67 indicates an improvement of two thirds of a letter grade above the average (e.g., going from a B- to a B+), and a *b*-value of 1.00 indicates an improvement of an entire letter grade above the average (e.g., going from a B- to an A-). Meanwhile, negative *b*-values indicate decreases in the letter grade. As letter grades are assigned in 0.33 increments, the findings were put in this context. The *b*-values of the attendance count variables can be easily compared to the other independent variables (IVs) by examining 10 units of a time. For instance, since the *b*-value for absences in LA Q1 is -0.03, then 10 absences results in a decrease of about one-third of a letter grade.

LA letter grades. The overall model was able to explain over 60% of the total variance of LA letter grades in Q1, Q2 and Q3. Demographics had a medium ES on LA letter grades, attendance had a small ES, and non-attendance comments had a large ES. The ΔR^2, *b*, *SE b*, and β values for the LA Q1 letter grades are presented in Table 12. The comparison between steps 1 and 4 reveals that after accounting for the non-attendance report card comments, the effect of gender decreased by approximately 60% and the effect of ethnicity decreased by approximately 40%. Even after controlling for report card comments, females and White students still performed better than male and Black students.

After controlling for demographics, grade level and attendance, students who received a positive self-management, positive other SEL, positive preparedness, or positive academic comment received on average about a third of a letter grade higher than the average student (e.g., going from a B- to a B); meanwhile, students who received a negative self-management, negative preparedness, or negative academic comment received on average between two-thirds to a full letter grade lower than the average

student (e.g., going from a B- to a C or C-). Thus, negative comments had two to three times greater of an effect on letter grades than positive ones. Notably, the ambiguous improvement comment *b*-value was always negative (i.e., indicating worse outcomes). Thus, students who received an ambiguous improvement comment received on average two-thirds of a letter grade lower than the average student (e.g., going from a B- to a C).

Overall, the results for Q2 and Q3 for LA letter grades were highly similar to those for Q1, with five exceptions. First, lunch status was significant in Q2 (b = -0.26, -0.24, and -0.18 in steps 1, 2 and 3, respectively), indicating that those qualifying for a free or reduced lunch received about one-sixth of a letter grade lower. Second, the negative attendance comment was stronger in Q2 (b = -1.01 and -0.85 in steps 2 and 3, respectively). Third, the positive academics comment was not significant in Q2. Fourth, the negative other SEL comment was significant in Q3 (b = -1.01), indicating a full letter grade decrease for those students. Notably, this was the only time its frequency was greater than 0.5% (see Table 5). Fifth, the ambiguous improvement comment was weaker in Q2 (b = -0.54) and Q3 (b = -0.51).

Likewise, overall the results for Q4 for LA letter grades were similar to those in Q1, with five exceptions. First, the ES for the comments step was weaker ($\Delta R^2 = .18$, which is a medium ES), which means that less total variance was explained ($R^2 = .45$). Second, the effect of gender and ethnicity only decreased by 30% and 25%, respectively, by step 4. Third, the negative attendance comment was stronger (b = -0.66). Fourth, the positive self-management (b = 0.47) and the positive other SEL (b = 0.54) comments were stronger. Fifth, the negative academics comment was not significant. One possible

reason for these first two exceptions is that fewer comments were received in Q4 than the other quarters (see Table 1 and Table 5).

MA letter grades. The overall model was able to explain about 50% of the total variance of MA letter grades in Q1, Q2 and Q3. Demographics had a medium ES on MA letter grades, grade level had a small ES, attendance had a small ES, and non-attendance comments had a large ES. The ΔR^2, *b*, *SE b*, and β values for MA Q1 letter grades are presented in Table 13. Overall, the results for Q2 and Q3 for MA letter grades were highly similar to those for Q1, except as noted two paragraphs below. The *b*-values for MA are interpreted the same as for LA. The comparison between the steps reveals that after accounting for the non-attendance report card comments, the effect of gender decreased by approximately 85% and the effect of ethnicity decreased by approximately 50%. This resulted in gender no longer having a significant effect after controlling for non-attendance comments. Nevertheless, White students still performed better than Black students, even after controlling for report card comments.

After controlling for demographics, grade level and attendance, students who received positive other SEL or positive preparedness comments received on average about a third of a letter grade higher than the average student (e.g., going from a B- to a B) and students who received positive academic comments received about two-thirds of a letter grade higher than the average student (e.g., going from a B- to a B+). Meanwhile, students who received negative self-management, negative other SEL, or negative academic comments received about two-thirds of a letter grade lower than the average student (e.g., going from a B- to a C), and those who received a negative preparedness comment received about a full letter grader lower than the average student (e.g., going

from a B- to a C-). Thus, as with LA, negative comments usually had two to three times greater of an effect on letter grades than positive ones. Notably, the ambiguous improvement comment *b*-value was negative (indicating worse outcomes). Students who received an ambiguous improvement comment received between two-thirds and a full letter grade lower than the average student (e.g., going from a B- to a C or C-).

While overall Q2 and Q3 were similar to Q1 for MA, there were some differences. First, the effect of gender was only reduced by about 50% in Q2, but gender was still not significant in step 4. Second, gender was not significant at all in Q3. Third, the effect of ethnicity was only reduced by about 40% in both Q2 and Q3. Fourth, grade level was not significant in step 4 of Q2 and it was never significant in Q3. Fifth, there were some fluctuations in the strengths of the relationships.

In a parallel manner, while Q4 is similar to Q1 for MA, there are some differences for Q4. First, the comments step had a medium ES ($\Delta R^2 = .17$) on letter grades and grade level was not significant, which meant that the overall model was able to explain 42% of the total variance. Two, the effect of gender was only reduced by about 10%, which means that gender was significant in step 4. Three, the effect of ethnicity was only reduced by about 30% by the comments step. Four, there were some fluctuations in the strengths of the relationships.

Standardized test scores. Multiple regression analyses were performed with the standardized test scores being the dependent variables (DVs), where step 1 was demographics (i.e., gender, ethnicity, SES, and their interactions), step 2 was attendance (i.e., tardies, absences, and attendance comments), step 3 was the comments in Q3, and step 4 was the letter grades in Q3. The standardized test scores were centered around the

mean for each subject area separately. Only Grade 11 was analyzed for these analyses, as it is the only grade tested in high schools in the state at the time of the research, which means the sample sizes were much smaller than the letter grade ones.

Originally, all steps were entered using the forced entry method. However, the comments step was not significant for either LA or MA even though they accounted for about 4% of the total variance. As the degrees of freedom for the comment steps may have constrained the analyses ($df_{1LA} = 9$, $df_{2LA} = 218$; $df_{1MA} = 9$, $df_{2MA} = 183$), the regressions were re-run with the comments step using the stepwise selection method. The stepwise selection method first removes the greatest *p*-value from the set of IVs one at a time until no *p*-values are greater than .10; then, it adds the smallest *p*-value from the set of IVs that are less than .05 one at a time; then the removal and addition processes repeat until no new variables can be removed or added. In practice, the stepwise selection method in SPSS is the similar to the forward stepwise method, except that each addition goes through a removal test (Field, 2009).

LA HSPA. The overall model was able to explain 41% of the total variance of LA HSPA scores. Demographics had a large ES ($\Delta R^2 = .31$) on LA HSPA scores, attendance had a small ES ($\Delta R^2 = .04$), the comments step had a small ES ($\Delta R^2 = .02$), and the letter grade in Q3 had a small ES ($\Delta R^2 = .04$). Only the positive preparation comment of Q3 was significant, but this comment was not significant after adding in the letter grade of Q3 in the next step. Notably, a large portion of the variance was influenced by the demographic variables of gender and ethnicity, primarily the later.

MA HSPA. The overall model was able to explain 41% of the total variance of the MA HSPA scores. Demographics had a large ES ($\Delta R^2 = .28$) on MA HSPA scores,

attendance had a small ES ($\Delta R^2 = .04$), the comments step had a small ES ($\Delta R^2 = .03$), and the letter grade in Q3 had a small ES ($\Delta R^2 = .07$). Only the positive academics comment of Q3 was significant, but this comment was not significant after adding in the letter grade of Q3 in the next step. Notably, a large portion of the variance was influenced by the demographic variable of ethnicity.

H3c: SEL Skills Predict Future Academic Success.

Letter Grades. For predicting future academic success in the classroom, the most critical end point is the end of the year (i.e., Q4). The comments and letter grades were examined sequentially in steps in order to determine the contribution of each quarter. Additionally, each comment category was examined separately for each subject area (i.e., LA and MA) in order to minimize potential problems with degrees of freedom due to the overlapping in variances between letter grades in each quarter and the final quarter and the overlapping variances between the comments and the letter grades. Therefore, multiple regression analyses were performed with the Q4 letter grades being the dependent variables (DVs), where step 1 was demographics (i.e., gender, ethnicity, SES, and their interactions), step 2 was grade level, step 3 was the Q1 letter grade, step 4 was attendance (i.e., tardies, absences, and attendance comments for all quarters), step 5 was the Q1 comments, step 6 was the Q2 letter grades, step 7 was the Q2 comments, step 8 was the Q3 letter grades, step 9 was the Q3 comments, and step 10 was the Q4 comments.

In the end, none of the results for the comments was interpretable. Most comments were not significant and the few times they were significant, inconsistent and/or odd results were discovered. For instance, sometimes positive comments predicted

a decrease in Q4 letter grades and sometimes negative comments predicted an increase in Q4 letter grades; however, bivariate correlations between these comments and the Q4 letter grades were in the expected direction. This suggests that the multicollinearity was a severe problem with these analyses. Alternative approaches, such as examining the positive comments as a group and the negative comments as a group, were equally uninterpretable. Therefore, the results are not presented for parsimony.

Standardized test scores. Multiple regression analyses were performed with the standardized test scores being the dependent variables (DVs), where step 1 was demographics (i.e., gender, ethnicity, SES, and their interactions), step 2 was attendance (i.e., tardies, absences, and attendance comments), step 3 was the comments in Q1, step 4 was the comments in Q2, step 5 was the comments in Q3, step 6 was the letter grades in Q1, step 7 was the letter grades in Q2, and step 8 was the letter grades in Q3. The standardized test scores were centered around the mean for each subject area separately. Only Grade 11 was analyzed for these analyses, as it is the only grade tested in high schools in the state at the time of the research, which means the sample sizes were much smaller than the letter grade ones.

In order to provide the best ease of interpretation while dealing with the degrees of freedom constraints that arose during the concurrent analyses for standardized test scores, a two-stage process was run. In stage one the comments steps were entered using the stepwise selection method. This helped prevent problems with the degrees of freedom that would result from adding 9 variables at once with the forced entry method. The non-comment steps were entered using the forced entry method, as done previously. In stage two the model was re-run with all steps using the forced entry method, with one

exception: If a comment category was significant in at least one quarter, that comment category was put into the model for all three quarters. This allowed the ability to interpret the comment category as a theoretical construct instead of merely as an empirical variable.

LA HSPA. The stepwise selection process from stage one identified the following 5 comment category variables as significant in the listed order: Negative self-management in Q1, positive other SEL in Q1, negative preparation in Q2, positive preparation in Q2, and ambiguous improvement in Q2. The model was then re-run with these comment categories included in all quarters for stage 2. The overall model for stage two explained 47% of the total variance of LA HSPA scores. Demographics had a large ES on LA HSPA scores, attendance had a small ES, the comments in Q1 and Q2 each had a small ES, the comments in Q3 were not significant, and the letter grades in Q2 had less than a small ES. Follow-up analyses revealed that letter grades in Q1 were not significant due to the comments and that letter grades in Q3 were not significant due to the letter grades in Q2.

The ΔR^2, b, $SE\ b$, and β values for the model of LA HSPA scores are presented in Table 14. After controlling for the modifiable factors (i.e., attendance, comments and letter grades), the initial effects of gender and ethnicity decreased by about 20%. The comment steps for the positive other SEL, positive preparation, negative self-management, negative preparation, and ambiguous improvement comment categories were significant and able to explain 11% of the variance of LA HSPA scores, which is a small ES. However, after adding in the letter grades for Q2, none of the comment

categories were significant, which may be due to their high shared variance as explained in the concurrent analyses.

***MA HSPA*.** The stepwise selection process form stage one identified the following 6 comment category variables as significant in the listed order: Negative preparation in Q1, positive academics in Q1, ambiguous improvement in Q2, positive academics in Q3, ambiguous improvement in Q3 and negative self-management in Q3. The model was then re-run with these comment categories included in all quarters for stage 2. The overall model for stage two explained 64% of the total variance of MA HSPA scores. Demographics had a large ES on MA HSPA scores, attendance was not significant, the comments in Q1, Q2 and Q3 each had a small ES, the letter grades in Q1 had a medium ES, the letter grades in Q2 had a small ES, and the letter grades in Q3 had less than a small ES.

The ΔR^2, *b*, *SE b*, and β values for the model of MA HSPA scores are presented in Table 15. After controlling for the modifiable factors (i.e., attendance, comments and letter grades)the initial effects of ethnicity (the only significant demographic variable) was reduced by about 40%. The comment steps for the positive academics, negative preparation, and ambiguous improvement comment categories were able to explain 16% of the variance of MA HSPA scores, which is a medium ES. Before adding the letter grades, the positive academics category predicted better HSPA scores and the negative comment categories of preparation and self-management predicted worse HSPA scores.

However, after adding the letter grades the negative self-management category predicted better HSPA scores, which was unexpected, and the negative preparation category was no longer significant. Meanwhile, the ambiguous improvement category

predicted both better HSPA scores, which was unexpected, and worse HSPA scores. Additionally, the higher letter grades in Q3 predicted worse HSPA scores, which was unexpected. An examination of the bivariate correlations for the three unexpected findings suggested that they were spurious and due to multicolinearity with the letter grades.

Revisiting the Logic Model

The preceding analyses allow for a revisiting of the logic model of the components of the report card (see Figure 4). The below summary and numbers in the figure were taken from the ΔR^2 values from the corresponding multiple regression steps for Q1 for language/literature and math in Tables 8 through 15, which are representative of the relationships in Q2 through Q4. The only exception is the values between the academic grades, where the bivariate correlations from the overall results were squared to get a rough equivalent to ΔR^2. There is no effect size provided for the two comment components (i.e., a to c in Figure 4) because there is no equivalent ΔR^2 value for comparing them. As the ΔR^2 values come from different multiple regression analyses, the ΔR^2 steps occur at different parts of the multiple regression analyses, and the numbers are not a result of structural equation modeling (SEM), the summary should be interpreted for the general principals expressed rather than the exact details.

Nevertheless, we can use Figure 4 to determine a rough picture of the relative strengths of the relationships between the different components of the report card. The strongest relationships were between the report card comments and the academic letter grades, which were in the large ES range. The next strongest relationship was between letter grades and standardized test scores, which was in the small to medium ES range.

The relationship between attendance and other components was in the small ES range. Finally, the relationship between the comments and standardized achievement test scores was in the medium ES range before adding the letter grades, but the relationship was non-significant after adding the letter grades.

Discussion

This dissertation was the first study to systematically and empirically study how the behavioral ratings and comments sections of the report card relate to SEL, academic grades, and standardized tests in a manner consistent with SEL theory. Analyses of over 1,000 report cards from an ethnically diverse, large high school revealed that report card comments had a small effect size on attendance and a large effect size on current letter grades. These findings applied to both "subjective" (i.e., language) and "objective" (i.e., mathematics) academic areas and were able to partially explain academic disparities for gender and ethnicity. However, the comments were only partially able to assess the five core aspects of SEL (i.e., self-awareness, self-management, social awareness, relationship skills, and responsible decision-making; CASEL, 2005). Additionally, the comments were unable to explain standardized test scores or predict future letter grades after accounting for prior letter grades. Nonetheless, methodological limitations and multicollinearity probably contributed to these null findings. Furthermore, the concurrent findings of this study suggest that with reforms report card comments may one day function as a feasible, scalable assessment system for SEL.

Consistency With Existing Literature

This study had many consistencies with the existing literature on report cards and on SEL. First, negative comments were more strongly associated with lower letter grades

than positive comments were associated with higher letter grades (Friedman, et al., 1998). Second, Black students and male students received more negative comments, less positive comments, and worse letter grades than their White and female peers as found in the prior SEL-IP study (Kemp, et al., 2012). This is noteworthy as the previous SEL-IP study occurred in a different school district, with different report card comments, and with different grade levels (i.e., 6th-8th grades instead of 9th-12th grades like this study). Third, fewer SEL skills, which include self-control, were associated with greater attendance problems, which parallels the finding that greater self-control is associated with fewer absences (Duckworth & Seligman, 2005). Fourth, the presence and absence of SEL skills was associated with higher and lower academic grades concurrently (Durlak, et al., 2011). This last point is especially notable as the presence and absence of SEL skills and behaviors in this study was based upon a new measuring method (i.e., report card comments) and this new method is based directly upon the students' skills and behaviors within the classroom.

This study also had many consistencies with the existing literature on academic disparities. First, Black students received lower standardized test scores than their White peers in both "subjective" (i.e., language/literature, reading) and "objective" (i.e., math) subject areas (Perie, Moran, & Lutkus, 2005). Second, the findings add support to the strong influence of SES on formal assessments of academic outcomes. In the present study students qualifying for a free or reduced lunch had lower standardized test scores in language/literature and math, which is congruent with the finding that there are differences between the 10th percentile and 90th percentile for SES (Reardon, 2011) and the finding that there are differences associated with parents' income, education and

occupation (Orr, 2003). Third, male students had lower standardized test scores than their female peers in the "subjective" area of language/literature, which parallels the finding that male students had lower standardized test scores than their female peers in the "subjective" area of reading (Perie, et al., 2005).

The findings regarding SES and ethnicity deserve further comment. It is critical to remember that all of the students in this study qualifying for a free or reduced lunch were Black, which means that SES was profoundly confounded with ethnicity in this study. This suggests that Black and White students at this school were coming from meaningfully different SES backgrounds, which was probably not fully accounted for by the free or reduced lunch status. Other studies have found that SES, when examined in greater detail than lunch status, has a much stronger effect on academic achievement than ethnicity (Orr, 2003; Reardon, 2011). Therefore, the findings regarding ethnicity in this study may really be due to unmeasured SES factors (Perie, et al., 2005).

Inconsistency With Existing Literature

Nevertheless, there were inconsistencies with the existing literature on SEL and academic discrepancies. First, the report card comments representing SEL were not associated with gains in letter grades or standardized test scores. This appears to be in contrast to the recent meta-analysis, which showed that universal school-based SEL interventions lead to an 11% gain in both letter grades and standardized test scores (Durlak, et al., 2011). However, the meta-analysis examined how different interventions (SEL vs. control) connected to future academic achievements. This study examined how different individuals' SEL skill levels and behaviors (positive or negative) connected to future academic achievements. As research has found that it takes 3-5 years for positive

changes to be noticeable and stable after implementing an SEL program (Dalton, et al., 2007), it may take more than one academic year for SEL skills within individuals to lead to improvements in their letter grades. This may be especially true when accounting for their previous letter grades and comparing them to students with equivalent prior letter grades.

Second, no differences were found on math academic achievement between the genders, whereas other studies have found that male students score higher than female students on math academic achievement in the high school years (Perie, et al., 2005). If the lack of replication for this study is due to the behaviors of teachers or parents, then this is good. However, the lack of replication may simply be due to the overwhelming influence of ethnicity in this study.

Commentary on the Different Types of Report Card Comment Categories

Some commentary on how the different types of comments affected the results is warranted. Notably, the below findings apply to both the "subjective" (i.e., lanaguage/literature) and the "objective" (i.e., math) classes.

First, negative comments were found to be more strongly associated with lower letter grades than positive comments were with higher letter grades. Theoretically, it is logical that a dearth of SEL skills can greatly interfere with acquiring academic skills and knowledge in school, even up to the point of failing, despite a strong positive learning environment and/or high cognitive intelligence. On the other hand, there is presumably an upper limit of how much an abundance of SEL skills can compensate for a poor learning environment and/or low cognitive intelligence.

Second, there was no meaningful difference between the SEL comment categories (i.e., self-management and other SEL) and the non-SEL comment categories (i.e., academics and preparedness). One explanation is that there are many paths to academic success and that cognitive intelligence, external events, and non-SEL skills, such as being prepared, also play key roles in academics. Another possibility is that a more thorough assessment of all of the categories may have led to meaningful differences between the categories.

A third alternative is that the comment categories were not actual assessments of their labels. Two hypothesis fall under this "false labeling" hypothesis. The "justification" hypothesis states that teachers felt compelled to justify letter grades that were low or high and did so through the comments. Meanwhile, the "reward/punishment" hypothesis states that students were given higher grades than earned for good behavior (e.g., positive self-management) and lower grades than earned for bad behavior (e.g., negative self-management).

The above "false labeling" hypothesis is unlikely, however. The key school informant said that teachers took the comments seriously. Further, the false labeling hypothesis does not explain why comments were associated with differences in attendance problems. Moreover, with this hypothesis teachers should be the most motivated to provide comments in the final marking period to justify their grades; but the exact opposite occurred with many more students receiving no comments in the fourth quarter than the earlier three quarters.

Limitations

The most important limitation of this study is that the report card structure only allowed teachers to assign up to two comments per quarter per class to a student. This severely limited the psychometrics that could be applied and the amount of information about a student's SEL and non-SEL skills and behaviors, especially since the comment category was either present or absent. While it is unknown if these results would generalize to schools that allow more than two comments per quarter or use other response formats, many schools, especially high schools, restrict report card comments ratings in the same fashion as the school studied here (Friedman & Frisbie, 1995).

Another limitation is that only one high school was examined with one set of report card comments. It is possible that different comments would result in different comment categories and different results. However, a study with the same report card structure is likely to run into similar methodological issues as this one. Furthermore, a prior study of a middle school found that only 52% of students received two comments when there was a two comment limit (Friedman, et al., 1998), so it is unlikely that other schools with similar report card structures provide a much greater frequency of comments.

Other important limitations are that the comments from this study did not represent all aspects of the CASEL 5 core competencies of SEL, that the same raters assigned the letter grades and report card comments at the same time, that there were no psychometrically accepted behavioral scales to compare to the report card comment categories, and that the data did not allow for teacher or class level analyses. However, it should be noted that teachers frequently complete measures in other studies, so that

limitation is less of an issue than the lack of external validations on the comment categories.

Future Research

Future research on the SEL-IP should maintain some of the aspects of this study while compensating for many of the above limitations. The theory and logic models do not require any major changes and the use of real attendance counts, real letter grades and real standardized test scores should be maintained. Two of the smaller changes needed for prospective studies are the inclusion of psychometrically accepted scales to check the validity of the report card comment categories and the and the ability to look at teacher level effects, as prior research has shown that the use of comments varies by teacher (Friedman, et al., 1998). Of course, there should be a series of studies done with schools of various grade configurations (i.e., elementary, middle and high schools) as well as demographic composition, in terms of ethnicity and SES.

However, the most important methodological change needed is a prospective, longitudinal empirical study of the effect of transforming the behavioral rating and comment section of the school report. The transformation should be done through an action-research paradigm to make sure that comments are useful for the researchers', school's, and teachers' purposes. Action-research refers to a process in which the investigation is systematic, participatory, reflective, and cyclic (Wadsworth, 1998).

The reform of this section of the report card is needed because methodological problems are likely to persist if the “choose two from a list” format continues. More variance in the response choices could be created by using a short set of questions on which all students are rated using a Likert scale. This would increase the probability of

predictive validity analyses. While some might view this as unfeasible, there are many short, psychometrically sound measures that assess SEL. For instance, the Devereaux Student Strengths Assessment Mini (DESSA-mini) is an 8-item screening tool for social-emotional competencies that has been standardized and norm-referenced on students in Kindergarten through 8th grade (Wilson-Ahlstrom, et al., 2011).

As the transformation of the report card might also increase the teachers' focus on SEL instruction, the study would ideally also include a measure of the overall school climate. The general school environment influences the individual's SEL skill development (Elias, 2009; Mattison & Aber, 2007). As the teacher's ratings of SEL skills may change as the norms within the school change, a pre/post assessment is needed at both the individual and the school level.

Implications for Current Practice

As a series of SEL-IP studies using the proposed changes will take multiple years to research, recommendations and implications for current practice with regards to the structure of report card comments and the academic discrepancies are given to assist schools in the present.

Structure of comments on the report cards. While future research on the SEL-IP is needed before report card comments can be ideally designed, this study can provide some guidelines. First, schools should reevaluate all of their comments and replace comments that are unhelpful or not actionable. For instance, ambiguous comments are misleading because they appear to be positive comments but they actual follow the same pattern as the negative comments. Ambiguous comments that focus on generic improvement are more like consolation prizes or backhanded compliments than a positive

sign of change. Meanwhile, the positive academics category is not actionable because it seems to describe an intrinsic trait of cognitive intelligence, which cannot be changed. So, even though this comment is positive, nothing can be done about it.

Comments should instead focus on skills, such as SEL skills like self-management, which can be taught rather than on intrinsic traits which cannot change. While this study was unable to find a link between the comments and future academic research, the research literature is clear that that teachable intrapersonal and interpersonal skills (i.e., SEL) can lead to as much improvements in academic achievement as pure academic interventions (CASEL, 2005; Durlak, et al., 2011), so there is no need to spend report card space talking about presumably stable traits. Schools should look to established theoretical paradigms, such as SEL (CASEL, 2005) or positive psychology (Duckworth, et al., 2011; Tough, 2011), to guide them in the formation of their report card comments.

Second, schools should consider revising the structure of their report cards. While it is common to limit report card comments to two per class (Friedman & Frisbie, 1995), this study has shown that this really limits the possible descriptions and constrains the assessment of students' SEL skills. One way to do this is to adapt the strategy currently (i.e., in 2011-2012) used at the same school as this study. According to the key school informant, parents can access their children's letter grades and behavioral comments on a secure website at any time (personal communication, March 16, 2012). These comments and letter grades can be updated at any time, instead of only four times a year with report cards. Moreover, the school now allows teachers to select comments from a list or write their own open-ended comments.

Academic discrepancies. This study joins the academic literature demonstrating the severe problem of academic discrepancies in this country between White students and Black students, between non-low SES students and low SES students, and between females and males. Demographic variables accounted for 3%-13% of the variance in attendance problems, 13%-20% of the variance in letter grades, and 28%-31% of the variance in standardized test scores in this study. These numbers fall within the small to medium, medium, and large effect size ranges, respectively, and leave little room for modifiable factors to make an impact. Therefore, policies that advocate exclusively for individuals and/or schools alone to overcome these social and economic influences are not realistic solutions (Rothstein, 2008). Instead, changes that promote greater social and economic equality as well as school reforms are needed (Rothstein, 2008).

While schools cannot solve the academic discrepancies problem alone, they should be more forthright about these disparities and seek community alliances to help alleviate them. Schools can hold in-service trainings for teachers and administrators to help them understand what factors they can control, such as the report card comments. Both this study and the previous SEL-IP study (Kemp, et al., 2012) found that Black students and male students received less positive comments and more negative comments than their White and female peers. While the data in this study does not allow us to determine why these differences occurred, schools should examine whether teacher biases led to inaccurate assessments or self-fulfilling prophecies (Kemp, et al., 2012). Schools should seek out the reason for these comment discrepancies and utilize evidence-based targeted interventions to ameliorate this problem. While many schools may not be

able to close the achievement gap for standardized test scores on their own, they should be able to close the achievement gap in report card comments.

Closing Remarks

Kazdin and Blase (2011) have argued that our current system of addressing mental health almost exclusively through individual psychotherapy will never be enough. Meanwhile, Durlak et al. (2011)'s meta-analysis has shown that universal school-based SEL interventions can make a meaningful impact on students' mental health, prosocial behavior and academic achievement. This study has demonstrated that report card comments are strongly linked to current academic letter grades, which suggests that report card comments may one day be able to assess SEL on a large-scale basis. This would allow for greater SEL dissemination and implementation. Future studies are needed in order to empirically understand the best way to do this. As policy makers, schools, and researchers work together to improve excellence in academics, they must also work together to increase equity among different demographic groups.

References

Academic, Social, and Emotional Learning Act of 2011, H. R. 2437, 112th Cong. (2011).

Bar-On, R. (2007). How important is it to educate people to be emotionally intelligent, and can it be done? In R. Bar-On, J. G. Mareen & M. J. Elias (Eds.), *Educating people tp be emotionally intelligent*. Westport, CT: Praeger Publishers.

Cohen, J. (1992). A Power Primer. *Pyschological Bulletin, 112*, 155-159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Collaborative for Academic, Social, and Emotional Learning. (2005) *Safe and Sound: An Educadtional Leader's Guide to Evidence-Based Social and Emotional Learning (SEL) Programs, Illinois Edition*. Chicago, IL: Author.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Dalton, J. H., Elias, M. J., & Wandersman, A. (2007). *Community psychology: Linking individuals and communities* (2nd ed.). Belmont, CA: Wadsworth Publishing.

DeFife, J. A. (2009). Effect Size Calculator and Conversions. [Excel] Retrieved May 4, 2011, from www.psychsystems.net/Manuals.

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2011, December 19). What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades. [Advanced online publication]. *Journal of Educational Psychology*. doi: 10.1037/a0026280

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*, 939-944.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*, 917-928.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405-432. doi: 10.1111/j.1467-8624.2010.01564.x

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J., et al. (2010). Youth risk behavior surveillance - United States, 2009. *Morbidity and Morality Weekly Report, 59*, SS-5.

Elias, M. J. (2007). From model implementation to sustainability: A multisite study of pathways to excellence in social-emotional learning and related school programs. In A. M. Blankenstein, P. D. Houston & R. W. Cole (Eds.), *Sustaining professional learning communities: The soul of educational leadership series* (pp. 59-95). Thousand Oaks, CA: Corwin Press.

Elias, M. J. (2009). Social-emotional and character development and academics as a dual focus of educational policy. *Educational Policy, 23*, 831-846.

Elias, M. J., & Moceri, D. C. (2012). *Developing social and emotional aspects of learning: The American experience*. Manuscript submitted for publication.

Elias, M. J., Wang, M. C., Weissberg, R. P., Zins, J. E., & Walberg, H. J. (2002). The other side of the report card: Student success depends on more than test scores. *American School Board Journal, 189*, 28-30.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE Publications Inc.

Fishman, D. B. (1999). *The Case for Pragmatic Psychology*. New York: New York University Press.

Fishman, D. B., Rego, S. A., & Muller, K. L. (2010). Behavioral theories of psychotherapy. In J. G. Norcross, G. R. VandernBos & D. K. Friedman (Eds.), *History of psychotherapy: Continuity and change* (2nd ed., pp. 101-140). Washington, DC: APA.

Friedman, S. J., & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational & Psychological Measurement, 55*, 5-26.

Friedman, S. J., Valde, G. A., & Obermeyer, B. J. (1998). Computerized report card comment menus: Teacher use and teacher parent perceptions. *Spectrum, 16*, 37-42.

Haahr, M. (2006). *Random.org: True random number service*, from http://www.random.org

Hall, G. E., & Hord, S. M. (2005). *Implementing Change: Patterns, Principles, and Potholes* (2nd ed.). Boston, MA: Allyn & Bacon.

Heckman, J. J., & Masterov, D. V. (2007, January). *The productivity argument for investing in young children*. Paper presented at the T.W. Schultz Award Lecture at the Annual Meeting of the Allied Social Sciences Association annual meeting, Chicago, IL. Retrived from http://jenni.uchicago.edu/human-inequality/papers/Heckman_final_all_wp_2007-03-22c_jsb.pdf

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172-177.

Horner, R. H., Todd, A. W., Lewis- Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The School-Wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions, 6*, 3-12.

Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness *Perspectives on Psychological Science, 6*, 21-37. doi: 10.1177/1745691610393527

Kemp, S. N., Moceri, D. C., & Elias, M. J. (2012). *Minority disproportionality in general education*. Manuscript in preparation.

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, L., Ormel, J., et al. (2009). The global burden of mental disorders: An update for the WHO World Mental Health (WMH) Surveys. *Epidemiologia e Psichiatria Sociale, 18*, 23-33.

Kuperminc, G. P., Leadbeater, B. J., & Blatt, S. J. (2001). School social climate and individual differences in vulternability to psychopathology among middle school students. *Journal of School Psychology, 39*, 141-159.

Lickona, T., & Davidson, M. L. (2003). *School as a Caring Community Profile-II (SCCP II).* Retrieved April 15, 2005, from http://www.cortland.edu/character/sccp-ii.htm

Mattison, E., & Aber, M. S. (2007). Closing the achievement gap: The association of racial climate with achievement and behavorial outcomes. *American Journal of Community Psychology, 40*, 1-12.

McCartney, K., & Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development, 71*, 173-180.

Moceri, D. C., Elias, M. J., Fishman, D. B., Pandina, R., & Reyes-Portillo, J. A. (2012). The urgency of doing: Assessing the system of sustainable implementation model via the Schools Implementing Towards Sustainability (SITS) scale. *Journal of Community Psychology, 40*, 501-519. doi: 10.1002/jcop.21477

Orr, A. J. (2003). Black-White differences in achievement: The importance of wealth. *Sociology of Education, 76*, 281-304.

Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics*. U.S. Department of Education,, Institute of Education Sciences, National Center for Education Statistics. Government Printing Office, Washington, DC.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. J. Duncan & R. Murnane (Eds.), *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*. New York, NY: Russell Sage Foundation.

Rothstein, R. (2008). Whose problem is poverty? *Educational Leardership, 65*, 8-13.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using Multivariate Statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Tough, P. (2011, September 14). What if the secret to success is failure? *New York Times* Retrieved October 4, 2011, from http://www.nytimes.com/2011/09/18/magazine/what-if-the-secret-to-success-is-failure.html?pagewanted=all

Wadsworth, Y. (1998). What is Participatory Action Research? *Action Research International*, Paper 2, Retrieved February 27, 2009 from http://www.scu.edu.au/schools/gcm/ar/ari/p-ywadsworth2098.html.

Wilson-Ahlstrom, A., Yohalem, N., DuBois, D., & Ji, P. (2011). From soft skills to hard data: Measuring youth program outcomes. *Forum for Youth Investment*.

Zins, J. E., Bloodworth, M. R., Weissberg, R. P., & Walberg, H. J. (2007). The scientific base linking social and emotional learning to school success. *Journal of Educational and Psychological Consultation, 17*, 191-210.

Zins, J. E., & Elias, M. J. (2006). Social and emotional learning. In G. G. Bear & K. M. Minke (Eds.), *Children's needs III* (pp. 1-13). Bethesda, MD: National Association of School Psychologists.

Appendix A

CASEL 5: Domains, Definitions, and Examples

Self-awareness: accurately assessing one's feelings, interests, values, and strengths; maintaining a well-grounded sense of self-confidence

- emotion recognition, identifying feelings
- forecasting emotion
- recognizing individual
- family strengths
- reflection on how one's behavior supports a caring community
- self-confidence
- interpreting emotions
- Understanding feeling words
- recognizing how emotions make our bodies feel
- self-worth
- understanding causes of emotions

Self-management: regulating one's emotions to handle stress, controlling impulses, and persevering in addressing challenges; expressing emotions appropriately; and setting and monitoring progress toward personal and academic goals

- establishing and monitoring class ground rules
- goals, setting/working on goals
- self-calming and control
- changing emotions
- setting/working on goals

- Calm-down strategies,
- expressing anger appropriately, anger management
- positive self-talk
- overcoming obstacles
- self-control
- managing emotions
- Impulse control
- emotion regulation
- expression

Social awareness: being able to take the perspective of and empathize with others; recognizing and appreciating individual and group similarities and differences; and recognizing and making best use of family, school, and community resources

- considering different points of view, recognizing feelings in others
- empathy
- appreciating diversity
- understanding facial verbal, and situational cues;
- intentionality
- perspective-taking
- using good manners
- appreciating differences
- Interpreting body language
- predicting and inferring others' emotions
- recognizing differences and commonalities

- recognizing feelings in others
- accurately assessing intentions
- interpreting social situations treating others fairly
- and being polite
- bystander awareness

Relationship skills: establishing and maintaining healthy and rewarding relationships based on cooperation; resisting inappropriate social pressure; preventing, managing, and resolving interpersonal conflict; and seeking help when needed;

- listening
- demonstrating respect and fairness
- being helpful
- being cooperative
- sharing
- taking turns
- group entry, Initiating positive relationships
- being a friend;
- handling teasing
- conflict resolution
- making up
- teamwork
- speaking skills,
- giving and receiving compliments
- reaching a consensus

- inclusion
- communication
- refusal skills;
- handling criticism,
- rejection and accusations;
- making amends

Responsible decision making: making decisions based on consideration of ethical standards, safety concerns, appropriate social norms, respect for others, and likely consequences of various actions; applying decision-making skills to academic and social situations; and contributing to the well-being of one's school and community

- social problem solving (setting social goals; evaluating possible consequences and solutions and outcomes)
- problem solving
- flexible thinking
- help seeking, asking for help
- safe and healthy choices
- Brainstorming,
- reaching a consensus
- cognitive, interpersonal, and group problem solving;
- flexible thinking
- fairness
- help seeking
- taking responsibility for oneself

- explaining one's reasoning,
- participating in group decision making and problem solving

Sources: CASEL (2005), Payton et al. (2000), Zins et al. (2007), and L. Dusenbury, vice president for research for CASEL (personal communication, August 16, 2011)

Appendix B

A Comparison of the Theoretical and and Logic Models

Comparing the partial theoretical (Figure 1) and logic models (Figure 3) reveals that (a) previous comments take the location of SEL skills, (b) attendance takes the location of attachment to school, (c) current comments takes the location of school behaviors, (d) previous academic grades takes the location of greater skills/knowledge acquired in school, and (e) current academic grades takes the location of academic grades. In both models, (f) standardized achievement test scores remains the same. Meanwhile, the link between (c) current comments and (b) attendance is no longer bidirectional; only the former leads to the later now. And, the pathway between (c) and (d) is eliminated as it is illogical for (c) current comments to lead to (d) previous academic grades. Finally, (g) events and (i) skills/knowledge that occur outside of school as well as (h) cognitive intelligence are removed from the model because they are not included on report cards.

It is important to note that there are many processes occurring within the theoretical model that are not accounted for by the logic model, as the report card being an incomplete assessment of students' lives. This is why some arrows were removed. For instance, school bonding (i.e., feeling physically safe and emotionally secure) is one way that attachment to school can influence school behaviors. As attendance (another aspect of attachment to school) is unlikely to influence report card comments (except when comments refer specifically to attendance problems), the pathway from school attachment to school behaviors was removed.

It is also important to note that some comments on the report card are likely to reflect issues unrelated to SEL skills and behaviors (e.g., a request for a parent teacher conference, or a praise for scholastic aptitude). Even though they do not reflect SEL skills, they may still have an

affect on current academic performance (i.e., e and f). In other words, SEL skills are just one category captured by report card comments.

Appendix C

Detailed Analyses of Total Comments and Letter Grades

Total Comments

LA. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 320.61$, $p < .000$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates ($\epsilon = .86$). The results show that the number of comments changed over time, $F(2.59, 2728.01) = 190.70$, $p < .000$, partial eta squared (η_p^2) = .15, which is a medium ES. Pairwise comparisons revealed that all quarters were significantly different from each other, with the exception of Q1 and Q3. However, difference contrasts revealed that the largest discrepancy is between Q4 and the other three quarters (Q1 to Q3), $F(1, 1053) = 702.46$, $p < .000$, $\eta_p^2 = .40$, which is a large ES, with Q4 having fewer comments than Q1 to Q3.

MA. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 326.61$, $p < .000$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates ($\epsilon = .83$). The results show that the number of comments changed over time, $F(2.50, 2120.08) = 205.55$, $p < .000$, $\eta_p^2 = .20$, which is a medium ES. Pairwise comparisons revealed that all quarters were significantly different from each other. However, difference contrasts revealed that the largest discrepancy is between Q4 and Q1 to Q3, $F(1, 848) = 652.49$, $p < .000$, , $\eta_p^2 = .44$, which is a large ES, with Q4 having fewer comments than Q1 to Q3.

Academic Letter Grades

LA. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 61.28$, $p < .000$; therefore, degrees of freedom were corrected using Greenhouse-

Geisser estimates (ϵ = .96). The results show that the letter grades changed over time, $F(2.88, 3032.52) = 12.89$, $p < .000$, partial eta squared (η_p^2) = .01. Pairwise comparisons revealed that all quarters were significantly different from each other, with the exceptions of Q1 and Q2 and of Q1 and Q3. However, difference contrasts revealed that the largest discrepancy is between Q4 and Q1 to Q3, $F(1, 1053) = 24.19$, $p < .000$, $\eta_p^2 = .02$, which is a small ES, with Q4 having lower letter grades than Q1 to Q3.

MA. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 73.05$, $p < .000$; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates (ϵ = .95). The results show that the letter grades changed over time, $F(2.84, 944.56 = 22.18$, $p < .000$, partial eta squared (η_p^2) = .03, which is a small ES. Pairwise comparisons revealed that Q1 was higher than each of the other three quarters. Helmert contrasts confirm that the largest discrepancy is between Q1 and Q2 to Q4, $F(1, 848) = 64.45$, $p < .000$, $\eta_p^2 = .07$, which is a small ES, with Q1 having higher letter grades than Q2 to Q4.

Appendix D

Repeated measure ANOVAs of Grade Level

Grade Level: LA

Repeated measure ANOVAs were performed with grade level as a between-subject factor for language/literature (LA) letter grades, positive comments, negative comments, and ambiguous comments. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. There was no significant difference by grade level for LA letter grades, F (3, 1050) = 0.16, p = .921, η_p^2 = .00. However, there was a significant grade level by time interaction, F (8.63, 3019.48) = 6.28, p < .000, η_p^2 = .02, which is a small ES. This interaction revealed that grades 9 and 11 were not equal in Q1, that all grade levels were equal in Q2, grades 9 and 12 did not equal grades 10 and 11 in Q3, and that grades 10 and 12 were not equal in Q4; the smallest mean difference was 0.153 and the largest mean difference was 0.217, both of which were closer to one-sixth of a letter grade difference than to a third of a letter grade difference.

Standardized Test Scores. No computations of grade level differences for LA HSPA scores was possible as the HSPA is taken in 11th grade.

Positive Comments. There was a significant difference by grade level for positive comments in LA across the four quarters, F (3, 1050) = 12.47, p < .000, η_p^2 = .03, which is a small ES. Those in 12th grade (M = 0.40, SE = 0.04) had the fewest positive comments, those in 9th grade (M = 0.67, SE = .03) had the most positive comments, and those in 10th (M = 0.51, SE = 0.03) and 11th (M = 0.56, SE = 0.03)

grades had equal amounts of positive comments. There was a significant grade level by time interaction, F (8.32, 2910.65) = 2.63, p = .006, η_p^2 = .01, which is less than a small ES. This interaction revealed that the above patterns occurred in every grade, except in Q4 where those in 11th grade also had fewer positive comments than those in 10th grade.

Negative Comments. There was no significant difference by grade level for negative comments in LA across the four quarters, F (3, 1050) = 1.62, p = .182, η_p^2 = .01. However, there was a significant grade level by time interaction, F (7.96, 2786.30) = 5.85, p < .000, η_p^2 = .02, which is a small ES. This interaction revealed that in Q2 those in 11th grade had more negative comments than those in 9th and 10th grades, that in Q3 those in 10th grade had more negative comments than those in 11th grade, and that in Q4 those in 11th grade had more negative comments than the other grade levels.

Ambiguous Comments. There was a significant difference by grade level for ambiguous comments in LA across the four quarters, F (3, 1050) = 5.17, p = .002, η_p^2 = .02, which is a small ES. Those in 9th (M = 0.06, SE = .01) and 10th grades (M = 0.06, SE = .01) had more ambiguous comments than those in 11th (M = 0.03, SE = 0.01) and 12th grades (M = 0.03, SE = 0.01).

Grade Level: MA

Repeated measure ANOVAs were performed with grade level as a between-subject factor for math (MA) letter grades, positive comments, negative comments, and ambiguous comments. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. There was a significant difference by grade level for MA letter grades across the four quarters, F (3, 845) = 4.60, p = .003, η_p^2 = .02, which is a small ES. Those in 9th grade (M = 2.22, SE = 0.06) received lower MA letter grades those in 10th (M = 2.54, SE = 0.07), 11th (M = 2.47, SE = 0.07), and 12th (M = 2.45, SE = 0.07) grades. There was also a significant grade level by time interaction, F (8.52, 2400.56) = 3.50, p < .000, η_p^2 = .00, which is less than a small ES. This interaction revealed that that same overall pattern happened across the four quarters with two exceptions: 1) in Q3 and in Q4 those in grade 10 received higher letter grades than those in grade 12, and 2) in Q4 those in 9th grade received equal letter grades to those in 12th grade.

Standardized Test Scores. No computations of grade level differences for MA HSPA scores was possible as the HSPA is taken in 11th grade.

Positive Comments. There was a significant difference by grade level for positive comments in MA across the four quarters, F (3, 845) = 3.78, p = .010, η_p^2 = .01, which is less than a small ES. Those in 9th grade (M = 0.51, SE = 0.02) had more positive comments than those in 11th grade; those in 10the grade (M = 0.52, SE = 0.03) had more positive comments than those in 11th (M = 0.40, SE = .03) and 12th (M = 0.45, SE = 0.03) grades. There was also a significant interaction between grade level and time, F (7.37, 2075.47) = 3.45, p = .001, η_p^2 = .01, which is less than a small ES. This interaction revealed that in Q1 those in 9th and 10th grades had more positive comments than those in 11th grade, that in Q2 those in 9th grade had more positive comments than those in 11th grade, that in Q3 those in 9th and 10th grades had more positive comments than those in 11th grade, and that in Q4 those in 10th and 12th grades had more positive comments than those in 9th grades.

Negative Comments. There was a significant difference by grade level for negative comments in MA across the four quarters, F (3, 845) = 7.21, $p < .000$, $\eta_p^2 = .03$, which is a small ES. Those in 9th grade ($M = 0.38$, $SE = 0.03$) had more negative comments than those in 10th ($M = 0.23$, $SE = 0.03$), 11th ($M = 0.23$, $SE = 0.03$), and 12th ($M = 0.27$, $SE = 0.03$) grades. There was also a significant interaction between grade level and time, F (8.07, 2272. 99) = 1.99, $p = .043$, $\eta_p^2 = .01$, which is less than a small ES. This interaction revealed that the same pattern across all quarters, with the exception that in Q3 those in 9th grade had the same amount of negative comments as those in 12th grade and that in Q3 those in 12th grade had more negative comments than those in 11th grade.

Ambiguous Comments. There was a significant difference by grade level for ambiguous comments in MA across the four comments, F (3, 845) = 7.27, $p < .000$, $\eta_p^2 = .03$, which is a small ES. Those in 9th grade ($M = 0.13$, $SE = .01$) had more ambiguous comments than those in 10th ($M = 0.07$, $SE = 0.01$), 11th ($M = 0.07$, $SE = 0.01$), and 12th ($M = 0.07$, $SE = 0.01$) grades. There was no significant grade level by time interaction, F (8.57, 2414.38) = 0.49, $p = .877$, $\eta_p^2 = .00$, which is a small ES.

Appendix E

Repeated measure ANOVAs of Ethnicity

Ethnicity: LA

Repeated measure ANOVAs were performed with ethnicity as a between-subject factor for language/literature (LA) letter grades, positive comments, negative comments, and ambiguous comments. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. Black students (M = 2.12, standard error [SE] = 0.04) received lower LA letter grades across the four quarters than White students (M = 3.15, SE = 0.04), F (1, 1052) = 309.05, p < .000, η_p^2 = .23, which is a medium ES. There was no ethnicity by time interaction, F (2.88, 3029.59) = 0.37, p = .77, η_p^2 = .00.

Standardized Test Scores. Black students (M = 221.06, SD = 20.32) received lower LA HSPA scores than White students (M = 245.13, SD = 16.16), t(230.89) = -10.07, p < .000, g = -1.29, which is a large ES.

Positive Comments. Black students (M = 0.42, SE = 0.02) received fewer positive comments in LA across the four quarters than White students (M = 0.72, SE = 0.02), F (1, 1052) = 87.19, p < .000, η_p^2 = .08, which is a small ES. There was a significant ethnicity by time interaction, F (2.77, 2912.56) = 10.20, p < .000, η_p^2 = .01, which is less than a small ES. This interaction revealed that the mean difference was only about half as large in Q4 as it was in Q1 through Q3.

Negative Comments. Black students (M = 0.51, SE = 0.02) received more negative comments in LA across the four quarters than White students (M = 0.25, SE =

.02), F (1, 1052) = 73.49, $p < .000$, $\eta_p^2 = .07$, which is a small ES. There was a significant ethnicity by time interaction, F (2.66, 2797.86) = 3.04, $p = .03$, $\eta_p^2 = .00$, which is less than a small ES. This interaction revealed that the mean difference was only about half as large in Q4 as it was in Q1 through Q3.

Ambiguous Comments. Black students ($M = .06$, $SE = .01$) received more ambiguous comments in LA across three of the four quarters than White students ($M = .03$, $SE = .01$), F (1, 1052) = 14.06, $p < .000$, $\eta_p^2 = .01$, which is a less than a small ES. There was a significant ethnicity by time interaction, F (2.62, 2752.96) = 5.15, $p = .003$, $\eta_p^2 = .01$, which is less than a small ES. This interaction revealed that the mean differences were significant in Q1 through Q3, albeit to different amounts (min = 0.03; max = 0.05), but not significant in Q4.

Ethnicity: MA

Repeated measure ANOVAs were performed with ethnicity as a between-subject factor for math (MA) letter grades, positive comments, negative comments, and ambiguous comments. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. Black students ($M = 2.10$, $SE = 0.04$) received lower MA letter grades across the four quarters than White students ($M = 2.88$, $SE = 0.05$), F (1, 847) = 141.80, $p < .000$, $\eta_p^2 = .14$, which is a medium ES. There was no ethnicity by time interaction, F (2.84, 2403.01) = 0.86, $p = .455$.

Standardized Test Scores. Black students (M = 215.50, SD = 27.30) received lower MA HSPA scores than White students (M = 247.44, SD = 17.89), $t(181.40) = -9.80$, $p < .000$, $g = -1.30$, which is a large ES.

Positive Comments. Black students (M = 0.41, SE = 0.02) received fewer positive comments in MA across the four quarters than White students (M = 0.57, SE = 0.02), F (1, 847) = 32.20, $p < .000$, $\eta_p^2 = .04$, which is small ES. There was a significant ethnicity by time interaction, F (2.45, 2075.32) = 5.63, $p = .002$, $\eta_p^2 = .01$, which is less than a small ES. This interaction revealed that the mean differences were significant in all quarters albeit to different amounts (min = 0.07; max = 0.26).

Negative Comments. Black students (M = 0.39, SE = 0.02) received more negative comments in MA across the four quarters than White students (M = 0.13, SE = 0.02), F (1, 847) = 88.17, $p < .000$, $\eta_p^2 = .09$, which is a small ES. There was no ethnicity by time interaction, F (2.69, 2278.22) = 2.38, $p = .07$, $\eta_p^2 = .00$.

Ambiguous Comments. Black students (M = 0.10, SE = .01) received more ambiguous comments in MA across the four quarters than White students (M = 0.06, SE = .01), F (1, 847) = 11.38, $p = .001$, $\eta_p^2 = .01$, which is less than a small ES. There was no ethnicity by time interaction, F (2.858, 2421.098) = 1.04, $p = .371$, $\eta_p^2 = .00$.

Appendix F

Repeated measure ANOVAs of Gender

Gender: LA

Repeated measure ANOVAs were performed with gender as a between-subject factor for language/literature (LA) letter grades, positive comments, negative comments, and ambiguous comments. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. Male students (M = 2.37, SE = 0.05) received lower LA letter grades across the four quarters than female students (M = 2.73, SE = .05), F (1, 1052) = 30.01, $p < .000$, $\eta_p^2 = .03$, which is small ES. There was a significant gender by time interaction, F (2.88, 3029.36) = 2.98, $p = .032$, $\eta_p^2 = .00$, which is less than a small ES. This interaction revealed that the mean differences were significant in all quarters albeit to different amounts (min = 0.29; max = 0.43).

Standardized Test Scores. Male students (M = 227.73, SD = 21.12) received lower LA HSPA scores than female students (M = 235.06 = 22.46), t(231) = -2.56, g = -0.34, which is a small ES.

Positive Comments. Male students (M = 0.47, SE = 0.02) received fewer positive comments in LA across the four quarters than female students (M = 0.62, SE = 0.02), F (1, 1052) = 19.96, $p < .000$, $\eta_p^2 = .02$, which is a small ES. There was no gender by time interaction, F (2.77, 2918.49) = 1.27, $p = .285$, $\eta_p^2 = .00$.

Negative Comments. Male students (M = 0.48, SE = 0.02) received more negative comments in LA across the four quarters than female students (M = 0.32, SE =

0.02), F (1, 1052) = 28.77, $p < .000$, $\eta_p^2 = .03$, which is a small ES. There was a significant gender by time interaction, F (2.67, 2796.00) = 3.12, $p = .031$, $\eta_p^2 = .00$, which is less than a small ES. This interaction revealed that the mean differences were significant in all quarters albeit to different amounts (min = 0.10; max = 0.22).

Ambiguous Comments. Male students ($M = 0.05$, $SE = .00$) received more ambiguous comments in LA across the four quarters than female students ($M = .04$, $SE = .00$), F (1, 1052) = 4.21, $p = .040$, $\eta_p^2 = .00$, which is less than a small ES. There was no gender by time interaction, F (2.62, 2756.55) = 0.54, $p = .633$, $\eta_p^2 = .00$.

Gender: MA

Repeated measure ANOVAs were performed with gender as a between-subject factor for math (MA) letter grades, positive comments, negative comments, and ambiguous comments. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. Male students ($M = 2.30$, $SE = 0.05$) received lower MA letter grades across the four quarters than female students ($M = 2.49$, $SE = .05$), F (1, 847) = 7.74, $p = .006$, $\eta_p^2 = .01$, which is less than a small ES. There was no gender by time interaction, F (2.84, 2404.14) = 0.16, $p = .916$, $\eta_p^2 = .00$.

Standardized Test Scores. There was no significant difference by gender for MA HSPA scores, t(190.32) = 0.10, $p = .92$, $g = .01$.

Positive Comments. There was no significant difference by gender for positive comments in MA across the four quarters, F (1, 847) = 0.89, $p = .346$, $\eta_p^2 = .00$; nor was

there a significant gender by time interaction, F (2.44, 2070.33) = 2.37, p = .082, η_p^2 = .00.

Negative Comments. Male students (M = 0.34, SD = .02) received more negative comments in LA across the four quarters than female students (M = 0.24, SE = 0.02), F (1, 847) = 12.85, p < .000, η_p^2 = .02, which is a small ES. There was no gender by time interaction, F (2.70, 2282.70) = 1.33, p = .263, η_p^2 = .00.

Ambiguous Comments. There was no significant difference by gender for ambiguous comments in MA across the four quarters, F (1, 847) = 0.00, p = .956, η_p^2 = .00; nor was there a significant gender by time interaction, F (2.86, 2420.66) = 0.65, p = .578, η_p^2 = .00.

Appendix G

Repeated measure ANOVAs of Lunch Status

Lunch Status: LA

Repeated measure ANOVAs were performed with lunch status as a between-subject factor for language/literature (LA) letter grades, positive comments, negative comments, and ambiguous comments. As all of the students in this sample qualifying for a free or reduced lunch were Black, White students were eliminated from these analyses. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. Students who qualified for a free or reduced lunch ($M = 1.96$, $SE = 0.07$) received lower LA letter grades across the four quarters than those who did not qualify ($M = 2.19$, $SE = 0.05$), $F\ (1, 601) = 5.81$, $p < .010$, $\eta_p^2 = .01$, which is a less than a small ES. There was no lunch status by time interaction, $F\ (2.86, 1716.43) = 0.58$, $p = .635$, $\eta_p^2 = .00$.

Standardized Test Scores. Students who qualified for a free or reduced lunch ($M = 212.73$, $SD = 20.50$) received lower LA HSPA scores than those who did not qualify ($M = 224.34$, $SD = 19.39$), $t(129) = 3.04$, $p < .003$, $g = -0.59$, which is a medium ES.

Positive Comments. There was no significant difference by lunch status for positive comments in LA across the four quarters, $F\ (1, 601) = 1.16$, $p < .282$, $\eta_p^2 = .00$; nor was there a significant lunch status by time interaction, $F\ (2.77, 1667.05) = 1.00$, $p = .386$, $\eta_p^2 = .00$.

Negative Comments. There was no significant difference by lunch status for negative comments in LA across the four quarters, $F\ (1, 601) = 0.52$, $p < .473$, $\eta_p^2 = .00$;

nor was there a significant lunch status by time interaction, F (2.59, 1553.35) = 0.13, p = .738, η_p^2 = .00.

Ambiguous Comments. There was no significant difference by lunch status for ambiguous comments in LA across the four quarters, F (1, 601) = 1.99, p = .159, η_p^2 = .00; nor was there a significant lunch status by time interaction, F (2.57, 1541.64) = 0.63, p = .573, η_p^2 = .00.

Lunch Status: MA

Repeated measure ANOVAs were performed with lunch status as a between-subject factor for math (MA) letter grades, positive comments, negative comments, and ambiguous comments. As all of the students in this sample qualifying for a free or reduced lunch were Black, White students were eliminated from these analyses. When Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates.

Letter Grades. There was no significant difference by lunch status for letter grades in MA across the four quarters, F (1, 517) = 0.00, p < .971, η_p^2 = .00; nor was there a significant lunch status by time interaction, F (1.78, 678.77) = 1.36, p = .255, η_p^2 = .00.

Standardized Test Scores. Students who qualified for a free or reduced lunch (M = 206.43, SD = 20.85) received lower MA HSPA scores than those who did not qualify (M = 219.14, SD = 28.80), t(91.20) = 2.79, p < .006, g = -0.47, which is a small ES.

Positive Comments. There was no significant difference by lunch status for positive comments in MA across the four quarters, F (1, 517) = 0.28, p = .600, η_p^2 = .00;

nor was there a significant lunch status by time interaction, F (2.49, 1288.98) = 0.32, p = .771, η_p^2 = .00.

Negative Comments. There was no significant difference by lunch status for negative comments in MA across the four quarters, F (1, 517) = 0.94, p < .334, η_p^2 = .03, which is a small ES. There was, however, a significant lunch status by time interaction, F (2.68, 1383.74) = 3.72, p = .014, η_p^2 = .01, which is less than a small ES. This interaction revealed that in Q1 students who qualified for a free or reduced lunch (M = 0.50, SE = 0.05, lower bound 95% confidence interval = 0.41, upper bound confidence interval = 0.60) received more negative comments than those who did not qualify (M = 0.34, SE = 0.04, lower bound 95% confidence interval = 0.27, upper bound 95% confidence interval = 0.42), whereas in the other three quarters there was no significant difference by lunch status.

Ambiguous Comments. There was no significant difference by lunch status for ambiguous comments in MA across the four quarters, F (1, 517) = 0.06, p = .804, η_p^2 = .00; nor was there a significant lunch status by time interaction, F (2.88, 1488.58) = 1.27, p = .285, η_p^2 = .00.

Table 1.
Participant Demographics and Averages

	LA: All Grades (n = 1054)		MA: All Grades (n = 849)		LA: 11th Grade (n = 233)		MA: 11th Grade (n = 195)	
Female (%)	542	(51.4)	455	(46.4)	123	(52.8)	112	(57.4)
African American (%)	603	(57.2)	519	(61.1)	131	(56.2)	129	(66.2)
Free/Reduced Lunch (%)	199	(18.9)	179	(21.1)	37	(15.9)	37	(19.0)
9th Grade (%)	321	(30.5)	260	(30.6)	-	-	-	-
10th Grade (%)	273	(25.9)	203	(23.9)	-	-	-	-
11th Grade (%)	233	(22.1)	196	(23.1)	233	(100)	195	(100)
12th Grade (%)	227	(21.5)	190	(22.4)	-	-	-	-
Q1 Grade Avg. (SD)	2.59	(1.16)	2.55	(1.08)	2.69	(1.16)	2.71	(1.03)
Q2 Grade Avg. (SD)	2.63	(1.20)	2.36	(1.14)	2.55	(1.19)	2.38	(1.15)
Q3 Grade Avg. (SD)	2.55	(1.17)	2.37	(1.14)	2.63	(1.11)	2.44	(1.12)
Q4 Grade Avg. (SD)	2.47	(1.25)	2.34	(1.21)	2.46	(1.27)	2.35	(1.24)
Final Mark (SD)	2.55	(1.10)	2.37	(1.03)	2.58	(1.08)	2.48	(0.99)
Tardy Avg. (SD)	5.16	(8.95)	6.16	(10.20)	6.58	(10.28)	6.46	(9.94)
Absence Avg. (SD)	13.08	(10.62)	13.25	(11.92)	13.13	(9.19)	14.22	(11.93)
HSPA Scale Score (SD)	-	-	-	-	231.60	(22.10)	226.31	(28.79)
Partially Proficient (%)	-	-	-	-	20	(8.6)	40	(20.5)
Proficient (%)	-	-	-	-	163	(70.0)	106	(54.4)
Advanced Proficient (%)	-	-	-	-	50	(21.5)	49	(25.1)
Total Comments for 2007-2008 Avg. (SD)	3.94	(2.00)	3.39	(1.81)	4.15	(2.25)	2.81	(1.68)
Q1 Com. Avg. (SD)	1.16	(0.73)	1.14	(0.63)	1.22	(0.76)	0.93	(0.61)
Q2 Com. Avg. (SD)	1.03	(0.70)	0.83	(0.68)	1.16	(0.75)	0.72	(0.70)
Q3 Com.Avg. (SD)	1.11	(0.66)	0.92	(0.67)	1.00	(0.76)	0.76	(0.74)

Q4 Com. Avg. (SD)	0.64	(0.68)	0.49	(0.65)	0.76	(0.70)	0.39	(0.59)
Q1 Zero Com. (%)	209	(19.8)	115	(13.5)	47	(20.2)	43	(22.1)
Q1 One Com. (%)	464	(44.0)	497	(58.5)	87	(37.3)	122	(62.6)
Q1 Two Com. (%)	381	(36.1)	237	(27.9)	99	(42.5)	30	(15.4)
Q2 Zero Com. (%)	246	(23.3)	281	(33.1)	50	(21.5)	83	(42.6)
Q2 One Com. (%)	534	(50.7)	429	(50.5)	96	(41.2)	84	(43.1)
Q2 Two Com. (%)	274	(26.0)	139	(16.4)	87	(37.3)	28	(14.4)
Q3 Zero Com. (%)	177	(16.8)	226	(26.6)	67	(28.8)	81	(41.5)
Q3 One Com. (%)	585	(55.5)	462	(54.4)	99	(42.5)	79	(40.5)
Q3 Two Com. (%)	292	(27.7)	161	(19.0)	67	(28.8)	35	(17.9)
Q4 Zero Com. (%)	495	(47.0)	505	(59.5)	91	(39.1)	130	(66.7)
Q4 One Com. (%)	440	(41.7)	272	(32.0)	106	(45.5)	54	(27.7)
Q4 Two Com. (%)	119	(11.3)	72	(8.5)	36	(15.5)	11	(5.6)

Notes: LA is language/literature classes. MA is math classes. All free/reduced lunch participants are African American/Black. The Grade Avg. for academic quarter one (Q1) to quarter four (Q4) is the letter grade on a 4-point scale (0 to 4 = F to A), where a 2.33 corresponds to a C+ and a 2.67 corresponds to a B-. The HSPA is a scaled score between 100 and 300. The maximum number of Total Comments for 2007-2008 is 8, as the maximum per semester is 2. The GPA is the average of the four quarters. The Final Mark is the final grade the student received for the official school records. The sample size for each row is the same as the column heading, unless otherwise specified. For Final Mark, the sample size is 1039, 829, 229, and 191 for the four columns respectively. For both Absence and Tardy, the sample size is 1053, 848, 232, and 195 for the four columns respectively.

Table 2.
List of 25 Report Card Comments

Comment ID	Comment
50	Dmstrts academic excellence
51	Demonstrates critical thinking
53	Shows improvement
53	Prepared & exhibits interest
54	Contributes to class
55	Inconsistent in effort, etc.
56	Needs to bring mtrls/assign
57	Often absent or late
58	Needs to seek help
59	Tests and quizzes
A	Is interested,eager to learn.
B	Constructive influence in cls.
C	Satisfactory effort attitudes.
D	Tries hard finds sub. diffclt.
E	Shows degree of aptitude.
F	Shows a lack of effort.
G	Interferes w/class - friction.
H	Is improving.
I	Grade affected by absences.
J	Grade affected by tardies.
K	Suggest parent-teacher confer.
L	Inconsistent in effort.
M	Not prepared with materials.
N	Discipline & disruptive behav.
O	Grade may affect level plcmnt.

Note: When the school gave us their report card data, the numbers and letters for comment IDs were included. The key school informant did not know why these specific labels existed.

Table 3.
Comment Categories: 3-Way Consensus

Self-Management	Other SEL	Non-SEL
"55" "Inconsistent in effort, etc." [2 agree self-mgmt, all 3 agree SEL]	"51" "Demonstrates critical thinking" [all 3 agreed other-SEL]	"50" "Dmstrts academic excellence" [all 3 agreed non-SEL]
"C" "Satisfactory effort attitudes." [2 agree self-mgmt, all 3 agree SEL]	"54" "Contributes to class" [all 3 agreed other-SEL]	"52" "Shows improvement" [1 said self-mgmt, 2 said non-SEL]
"F" "Shows a lack of effort." [2 agree self-mgmt, all 3 agree SEL]	"58" "Needs to seek help" [all 3 agreed other-SEL]	"53" "Prepared & exhibits interest" [2 said SEL (one self-mgmt), 1 said non-SEL]
"L" "Inconsistent in effort." [2 agree self-mgmt, all 3 agree SEL]	"B" "Constructive influence in cls." [all 3 agreed other-SEL]	"56" "Needs to bring mtrls/assign" [all 3 agreed non-SEL]
"N" "Discipline & disruptive behav." [all 3 agree self-mgmt]	"G" "Interferes w/class - friction." [all 3 agreed other-SEL]	"57" "Often absent or late" [all 3 agree non-SEL]
		"59" "Tests and quizzes" [all 3 agreed non-SEL]
		"A" "Is interested,eager to learn." [2 said self-mgmt, 1 said non-SEL]
		"D" "Tries hard finds sub. diffclt." [2 said self-mgt, 1 said non-SEL]
		"E" "Shows degree of aptitude." [all 3 agreed non-SEL]
		"H" "Is improving." [1 said self-mgmt, 2 said non-SEL]
		"I" "Grade affected by absences." [all 3 agreed non-SEL]
		"J" "Grade affected by tardies." [all 3 agreed non-SEL]
		"K" "Suggest parent-teacher confer." [all 3 agreed non-SEL]
		"M" "Not prepared with materials." [all 3 agreed non-SEL]
		"O" "Grade may affect level plcmnt." [all 3 agreed non-SEL]

Note: Three sources contributed to this table. The first is the consensus of two SEL experienced undergraduate research assistants. The second and third are two school psychologists who are SEL experts. The following rule system was applied to reach consensus: 1) If two sources put a comment in the self-management category and the third put it in the other-SEL category, it was labeled as self-management; 2) If all three sources put the comment in an SEL category, it was labeled as other-SEL; 3) If at least one source put the comment in the non-SEL category, it was labeled as non-SEL (usually

if one did, all three did). There was an 80% agreement across the three sources when collapsing the two SEL categories (i.e., both self-management and other SEL) and comparing that to the non-SEL category.

Table 4.
Comment Categories: 10 Final Categories

	Self-Management	**Other SEL**	**Non-SEL: Academic/Extremes/Other**	**Non-SEL: Attendance**	**Non-SEL: Improvement/ Ambivalence**	**Non-SEL: Preparedness/Interest**
Positive	"C" "Satisfactory effort attitudes." [2 agree self-mgmt, all 3 agree SEL]	"51" "Demonstrates critical thinking" [all 3 agreed other-SEL]	"50" "Dmstrts academic excellence" [all 3 agreed non-SEL]			"53" "Prepared & exhibits interest" [2 said SEL (one self-mgmt), 1 said non-SEL]
		"54" "Contributes to class" [all 3 agreed other-SEL]	"E" "Shows degree of aptitude." [all 3 agreed non-SEL]			"A" "Is interested, eager to learn." [2 said self-mgmt, 1 said non-SEL]
		"B" "Constructive influence in cls." [all 3 agreed other-SEL]				
Negative	"55" "Inconsistent in effort, etc." [2 agree self-mgmt, all 3	"58" "Needs to seek help" [all 3 agreed other-SEL]	"56" "Needs to bring mtrls/assign" [all 3 agreed non-SEL]	"57" "Often absent or late" [all 3 agree non-SEL]	"52" "Shows improvement" [1 said self-mgmt, 2 said non-SEL]	"M" "Not prepared with materials." [all 3 agreed non-SEL]

	agree SEL]					
	"F" "Shows a lack of effort." [2 agree self-mgmt, all 3 agree SEL]	"G" "Interferes w/class - friction." [all 3 agreed other-SEL]	"59" "Tests and quizzes" [all 3 agreed non-SEL]	"I" "Grade affected by absences." [all 3 agreed non-SEL]	"D" "Tries hard finds sub. diffclt." [2 said self-mgt, 1 said non-SEL]	
	"L" "Inconsistent in effort." [2 agree self-mgmt, all 3 agree SEL]		"K" "Suggest parent-teacher confer." [all 3 agreed non-SEL]	"J" "Grade affected by tardies." [all 3 agreed non-SEL]	"H" "Is improving." [1 said self-mgmt, 2 said non-SEL]	
	"N" "Discipline & disruptive behav." [all 3 agree self-mgmt]		"O" "Grade may affect level plcmnt." [all 3 agreed non-SEL]			

Table 5.
Frequency of Comment Categories for Language/Literature: All Grades

	Q1	Q2	Q3	Q4
Self-Management (Positive)	19.3	17.7	13.8	8.1
Self-Management (Negative)	18.8	20.2	24.8	11.7
Other SEL (Positive)	25.6	17.9	24.0	11.4
Other SEL (Negative)	0.3	0.3	0.6	0.1
Academic/Extremes/Other (Positive)	5.7	4.2	5.7	3.5
Academic/Extremes/Other (Negative)	8.4	8.0	7.4	6.8
Attendance (Negative)	3.5	3.7	3.6	2.6
Improvement (Ambivalent)	4.8	4.1	6.5	2.0
Preparedness/Interest (Positive)	18.1	15.7	14.4	9.2
Preparedness/Interest (Negative)	9.0	7.9	8.0	6.5

Notes: This table shows the percentage of students who received at least one comment for the comment category. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively. There were 1,054 students for this analysis.

Table 6.
Frequency of General Comment Categories for All Quarters: All Grades

	F		D		C		B		A	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Pos Com LA	0.2	1.0	0.6	2.1	6.7	11.3	41.1	53.3	38.6	49.7
Pos Com MA	0.0	1.5	1.9	6.2	21.2	25.8	39.0	48.6	27.4	34.8
Neg Com LA	21.9	29.1	23.5	31.1	26.3	32.6	15.9	20.3	0.6	3.2
Neg Com MA	23.1	34.0	22.4	37.0	19.4	37.7	5.9	16.4	0.0	2.8
Ambig Com LA	0.0	19.0	14.3	33.3	25.6	45.1	13.7	37.2	0.0	4.8
Ambig Com MA	4.2	24.4	29.2	33.8	31.1	45.8	9.1	22.9	0.0	5.2

Notes: This table shows the percentage of students who received at least one comment for the general comment categories for that particular letter grade (F through A). The percentages for each quarter were examined and the minimum and maximum values were entered in their respective cells. Letter grades include pluses and minuses. There were 1,054 students for language/literature (LA) and 849 for math (MA). Pos is positive. Neg is negative. Com is comments.

Table 7.
Frequency of Positive and Negative Comments for Language/Literature: All Grades

		Q1		Q2		Q3		Q4	
Pos	Avg. (SD)	0.70	(0.73)	0.56	(0.66)	0.58	(0.67)	0.33	(0.54)
	Zero Com (%)	487	(46.2)	563	(53.4)	549	(52.1)	741	(70.3)
	One Com (%)	394	(37.4)	388	(36.8)	394	(37.4)	278	(26.4)
	Two Com (%)	173	(16.4)	103	(09.8)	111	(26.4)	35	(03.3)
Neg	Avg. (SD)	0.41	(0.68)	0.42	(0.69)	0.46	(0.70)	0.29	(0.56)
	Zero Com (%)	739	(70.1)	735	(69.7)	695	(65.9)	803	(76.2)
	One Com (%)	196	(18.6)	194	(18.4)	234	(22.2)	195	(18.5)
	Two Com (%)	119	(11.3)	125	(11.9)	125	(11.9)	56	(05.3)
Ambig	Avg. (SD)	0.05	(0.21)	0.04	(0.20)	0.07	(0.25)	0.02	(0.14)
	Zero Com (%)	1003	(95.2)	1011	(95.9)	985	(93.5)	1033	(98.0)
	One or Two Com (%)	51	(04.8)	43	(04.1)	69	(06.5)	21	(02.0)

Notes: Total positive comments is abbreviated as Pos. Total negative comments is abbreviated as Neg. Comments is abbreviated as Com. Total ambiguous comments is abbreviated as Ambig. Very few students had more than one ambiguous comment, so having one or two comments were combined. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively. There were 1,054 students for this analysis

Table 8.
Multiple Regression for Tardies for Language/Literature: All Grades

	ΔR^2	*B*	*SE B*	β
Ethnicity	.06**	-1.94	0.37	-0.14**
Grade 9	.05**	-2.56	.50	-0.17**
Grade 10		-2.29	0.51	-0.15**
Grade 12		0.80	0.53	0.05
Neg Atnd Q1	.14**	5.54	1.01	0.15**
Neg Atnd Q2		2.30	1.09	0.65*
Neg Atnd Q3		8.14	1.00	0.22**
Neg Atnd Q4		3.21	1.32	0.08*
Neg Self-Mgmt Q1	.07**	1.71	0.49	0.10**
Neg Self-Mgmt Q2		1.75	0.51	0.10**
Neg Acad Q2		-2.22	0.71	-0.09**
Ambig Improv Q2		2.12	0.90	0.06*
Neg Prep Q3		1.65	0.67	0.07*
Neg Acad Q3		1.75	0.74	0.07*
Neg Self-Mgmt Q4		2.40	0.59	0.11**

Note: $R^2 = .31$ (ΔR^2s add up to .32 due to rounding); $n = 1053$, one participant did not have attendance data. For ethnicity, 0 = Black, 1 = White. For the grade levels, 0 = grade 11 and 1 = grade 9, 10 or 12. For comments, 0 = did not receive this comment, 1 = did receive this comment. Neg is negative. Ambig is ambiguous. Atnd is the attendance comments. Self-Mgmt is the self-management comment. Acad is the academics comment. Improv is the improvement comment. Prep is the preparation comment. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively.
* $p < .05$, ** $p < .01$

Table 9.
Multiple Regression for Absences for Language/Literature: All Grades

	ΔR^2	*B*	*SE B*	β
Gender	.04**	2.29	0.41	0.16**
Ethnicity		3.01	0.42	0.20**
Grade 9	.11**	-2.78	0.57	-0.17**
Grade 10		-1.28	0.58	-0.08*
Grade 12		3.67	0.61	0.21**
Neg Atnd Q1	.05**	1.85	1.14	0.05
Neg Atnd Q2		7.40	1.24	0.19**
Neg Atnd Q3		3.61	1.15	0.09**
Neg Atnd Q4		-2.23	1.51	-0.05
Neg Self-Mgmt Q1	.02**	2.58	0.54	0.14**
Pos Other SEL Q4		-1.42	0.65	-0.06*

Note: $R^2 = .22$; $n = 1053$, one participant did not have attendance data. For gender, 0 = male, 1 = female. For ethnicity, 0 = Black, 1 = White. For the grade levels, 0 = grade 11 and 1 = grade 9, 10 or 12. For comments, 0 = did not receive this comment, 1 = did receive this comment. Pos is positive. Neg is negative. Atnd is the attendance comments. Self-Mgmt is the self-management comment. Other SEL is the other SEL comments. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively.
* $p < .05$, ** $p < .01$

Table 10.
Multiple Regression for Tardies for Math: All Grades

	ΔR^2	*B*	*SE B*	β
Gender	.13**	-1.01	0.44	-0.07*
Ethnicity		-3.36	0.50	-0.22**
Free/Reduced Lunch		1.16	0.58	0.06*
Neg Atnd Q1	.15**	7.20	1.47	0.17**
Neg Atnd Q2		-4.88	2.08	-0.10*
Neg Atnd Q3		10.91	1.16	0.30**
Neg Atnd Q4		5.19	2.03	0.10*
Pos Self-Mgmt Q1	.06**	1.62	0.54	0.09**
Neg Other SEL Q1		-3.65	1.40	-0.08**
Neg Self-Mgmt Q2		2.27	0.67	0.11**
Neg Other SEL Q2		7.89	3.15	0.13*
Neg Prep Q2		3.75	0.92	0.12**
Neg Self-Mgmt Q3		1.73	0.76	0.07*
Ambig Improv Q3		1.78	0.69	0.07**
Neg Other SEL Q4		-9.82	3.78	-0.13**

Note: $R^2 = .34$; $n = 848$, one participant did not have attendance data. For gender, 0 = male, 1 = female. For ethnicity, 0 = Black, 1 = White. For free/reduced lunch, 0 = does not qualify for a free or reduced lunch, 1 = does qualify. For comments, 0 = did not receive this comment, 1 = did receive this comment. Pos is positive. Neg is negative. Ambig is ambiguous. Atnd is the attendance comments. Self-Mgmt is the self-management comment. Other SEL is the other SEL comment. Prep is the preparation comment. Improv is the improvement comment. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively.
* $p < .05$, ** $p < .01$

Table 11.
Multiple Regression for Absences for Math: All Grades

	ΔR^2	*B*	*SE B*	β
Gender	.03**	1.42	0.45	0.10**
Ethnicity		2.54	0.46	0.17**
Grade 9	.10**	-3.22	0.61	-0.20**
Grade 10		-1.41	0.65	-0.08*
Grade 12		3.26	0.66	0.19**
Neg Atnd Q1	.10**	-0.07	1.51	-0.00
Neg Atnd Q2		9.04	2.14	0.20**
Neg Atnd Q3		5.49	1.19	0.16**
Neg Atnd Q4		1.70	2.07	0.04

Note: $R^2 = .23$; $n = 848$, one participant did not have attendance data. For gender, 0 = male, 1 = female. For ethnicity, 0 = Black, 1 = White. For the grade levels, 0 = grade 11 and 1 = grade 9, 10 or 12. For comments, 0 = did not receive this comment, 1 = did receive this comment. Neg is negative. Atnd is the attendance comments. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively.
* $p < .05$, ** $p < .01$

Table 12.
Multiple Regression for Letter Grades for Language/Literature in Quarter 1: All Grades

		ΔR^2	*B*	*SE B*	β
Step 1	Gender	0.20**	0.31	0.06	0.13**
	Ethnicity		1.02	0.07	0.43**
Step 2	Gender	0.06**	0.37	0.06	0.16**
	Ethnicity		1.00	0.07	0.43**
	Tardies (count)		-0.02	0.01	-0.12**
	Absences (count)		-0.03	0.00	-0.18**
	Neg Atnd		-0.41	0.17	-0.07*
Step 3	Gender	0.37**	0.13	0.05	0.06**
	Ethnicity		0.60	0.05	0.26**
	Tardies (count)		-0.00	0.00	-0.02
	Absences (count)		-0.01	0.00	-0.09**
	Neg Atnd		-0.33	0.13	-0.05**
	Pos Self-Mgmt		0.27	0.06	0.09**
	Pos Other SEL		0.31	0.06	0.12**
	Pos Prep		0.41	0.06	0.14**
	Pos Acad		0.42	0.10	0.08**
	Neg Self-Mgmt		-0.90	0.06	-0.30**
	Neg Prep		-1.13	0.08	-0.28**
	Neg Acad		-0.83	0.08	-0.20**
	Ambig Improv		-0.81	0.10	-0.15**

Note: $R^2 = .64$; $n = 1053$, one participant did not have attendance data. For gender, 0 = male, 1 = female. For ethnicity, 0 = Black, 1 = White. For comments, 0 = did not receive this comment, 1 = did receive this comment. Pos is positive. Neg is negative. Ambig is ambiguous. Atnd is the attendance comments. Self-Mgmt is the self-management comment. Other SEL is the other SEL comment. Prep is the preparation comment. Acad is the academics comment. Improv is the improvement comment.
* $p < .05$, ** $p < .01$

Table 13.
Multiple Regression for Letter Grades for Math in Quarter 1: All Grades

		ΔR^2	*B*	*SE B*	β
Step 1	Gender	.13**	0.22	0.07	0.10**
	Ethnicity		0.78	0.07	0.35**
Step 2	Gender	.02**	0.21	0.07	0.10**
	Ethnicity		0.78	0.07	0.35**
	Grade 9		-0.41	0.09	-0.17**
	Grade 10		-0.18	0.10	-0.07
	Grade 12		-0.11	0.10	-0.04
Step 3	Gender	.11**	0.21	0.07	0.10**
	Ethnicity		0.65	0.07	0.29**
	Grade 9		-0.50	0.09	-0.22**
	Grade 10		-0.27	0.09	-0.11**
	Grade 12		-0.07	0.10	-0.03
	Tardies (count)		-0.03	0.01	-0.19**
	Absences (count)		-0.02	0.01	-0.16**
	Neg Atnd		-0.91	0.19	-0.15**
Step 4	Gender	.29**	0.03	0.05	0.02
	Ethnicity		0.35	0.06	0.16**
	Grade 9		-0.31	0.07	-0.13**
	Grade 10		-0.32	0.08	-0.13**
	Grade 12		-0.10	0.08	-0.04
	Tardies (count)		-0.02	0.00	-0.13**
	Absences (count)		-0.02	0.00	-0.10**
	Neg Atnd		-0.88	0.15	-0.14**
	Pos Other SEL		0.44	0.07	0.15**
	Pos Prep		0.34	0.06	0.15**
	Pos Acad		0.75	0.11	0.16**
	Neg Self-mgmt		-0.72	0.10	-0.21**
	Neg Other SEL		-0.74	0.17	-0.11**
	Neg Prep		-0.93	0.11	-0.23**
	Neg Acad		-0.58	0.13	-0.11**
	Ambig Improv		-0.83	0.10	-0.22**

Note: $R^2 = .55$; $n = 848$, one participant did not have attendance data. For gender, 0 = male, 1 = female. For ethnicity, 0 = Black, 1 = White. For the grade levels, 0 = grade 11 and 1 = grade 9, 10 or 12. For comments, 0 = did not receive this comment, 1 = did receive this comment. Pos is positive. Neg is negative. Ambig is ambiguous. Atnd is the attendance comments. Self-Mgmt is the self-management comment. Other SEL is the other SEL comment. Prep is the preparation comment. Acad is the academics comment. Improv is the improvement comment.
* $p < .05$, ** $p < .01$

Table 14.
Multiple Regression for Standardized Test Scores for Language/Literature: 11th Graders

		ΔR^2	*B*	*SE B*	β
Step 1	Gender	.31**	6.72	2.41	0.15**
	Ethnicity		23.77	2.43	0.54**
Step 2	Gender	.04**	6.89	2.34	0.16**
	Ethnicity		20.65	2.48	0.47**
	Tardies (count)		-0.65	0.16	-0.22**
Step 3	Gender	.06**	5.95	2.35	0.14*
	Ethnicity		19.68	2.42	0.44**
	Tardies (count)		-0.51	0.16	-0.18**
	Pos Other SEL Q1		7.20	3.14	0.13*
	Pos Prep Q1		-1.77	3.04	-0.03
	Neg Self-mgmt Q1		-7.68	3.27	-0.13*
	Neg Prep Q1		-7.06	4.24	-0.09
	Ambig Improv Q1		-8.80	7.90	-0.06
Step 4	Gender	.05**	5.94	2.35	0.14*
	Ethnicity		19.01	2.38	0.43**
	Tardies (count)		-0.42	0.16	-0.14*
	Pos Other SEL Q1		5.17	3.17	0.09
	Pos Prep Q1		-4.72	3.12	-0.09
	Neg Self-mgmt Q1		-5.80	3.30	-0.10
	Neg Prep Q1		-2.65	4.36	-0.03
	Ambig Improv Q1		-7.05	7.82	-0.05
	Pos Other SEL Q2		-1.32	3.49	-0.02
	Pos Prep Q2		6.82	3.12	0.13*
	Neg Self-mgmt Q2		-4.23	3.00	-0.09
	Neg Prep Q2		-9.19	4.05	-0.12*
	Ambig Improv Q2		-13.98	6.84	-0.11*
Step 5	Gender	.01	5.71	2.41	0.13*
	Ethnicity		19.10	2.39	0.43**
	Tardies (count)		-0.41	0.17	-0.14*
	Pos Other SEL Q1		5.40	3.21	0.10
	Pos Prep Q1		-4.76	3.18	-0.09
	Neg Self-mgmt Q1		-5.73	3.34	-0.10
	Neg Prep Q1		-2.58	4.60	-0.03
	Ambig Improv Q1		-9.59	8.17	-0.06
	Pos Other SEL Q2		-2.33	3.65	-0.04
	Pos Prep Q2		5.89	3.37	0.11
	Neg Self-mgmt Q2		-3.73	3.07	-0.08
	Neg Prep Q2		-9.15	4.27	-0.12*
	Ambig Improv Q2		-12.62	7.01	-0.10
	Pos Other SEL Q3		3.10	3.18	0.06
	Pos Prep Q3		0.97	3.20	0.02

	Neg Self-mgmt Q3		-0.77	3.06	-0.01
	Neg Prep Q3		3.61	4.54	0.05
	Ambig Improv Q3		-2.57	6.10	-0.02
Step 6	Gender	.01*	5.56	2.39	0.13*
	Ethnicity		18.38	2.40	0.41**
	Tardies (count)		-0.36	0.17	-0.12*
	Pos Other SEL Q1		4.83	3.19	0.09
	Pos Prep Q1		-4.30	3.16	-0.08
	Neg Self-mgmt Q1		-4.46	3.36	-0.08
	Neg Prep Q1		-1.43	4.59	-.02
	Ambig Improv Q1		-9.27	8.10	-0.06
	Pos Other SEL Q2		-3.47	3.66	-0.06
	Pos Prep Q2		4.73	3.39	0.09
	Neg Self-mgmt Q2		-0.24	3.44	-0.01
	Neg Prep Q2		-4.57	4.73	-0.06
	Ambig Improv Q2		-10.89	7.00	-0.09
	Pos Other SEL Q3		2.25	3.18	0.04
	Pos Prep Q3		0.96	3.17	0.02
	Neg Self-mgmt Q3		0.16	3.06	0.00
	Neg Prep Q3		3.36	4.50	0.04
	Ambig Improv Q3		-1.96	6.05	-0.02
	LA Letter Grades Q2		3.53	1.62	0.19*

Note: R^2 = .47 (ΔR^2s add up to .48 due to rounding); n = 232, one participant did not have attendance data. For gender, 0 = male, 1 = female. For ethnicity, 0 = Black, 1 = White. For comments, 0 = did not receive this comment, 1 = did receive this comment. Pos is positive. Neg is negative. Ambig is ambiguous. Atnd is the attendance comments. Self-Mgmt is the self-management comment. Other SEL is the other SEL comment. Prep is the preparation comment. Acad is the academics comment. Improv is the improvement comment. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively.

* $p < .05$, ** $p < .01$

Table 15.
Multiple Regression for Standardized Test Scores for Math: 11th Graders

		ΔR^2	*B*	*SE B*	β
Step 1	Ethnicity	.28**	31.94	3.71	0.53**
Step 2	Ethnicity	.08**	29.22	3.64	0.48**
	Pos Acad Q1		16.04	7.68	0.12*
	Neg Self-Mgmt Q1		-5.12	6.59	-0.05
	Neg Prep Q1		-23.87	7.94	-0.19**
	Ambig Improv Q1		-14.80	7.38	-0.12*
Step 3	Ethnicity	.04*	29.03	3.58	0.48**
	Pos Acad Q1		15.42	8.09	0.12
	Neg Self-Mgmt Q1		-2.37	6.68	-0.02
	Neg Prep Q1		-21.10	7.97	-0.17**
	Ambig Improv Q1		-8.49	7.52	-0.07
	Pos Acad Q2		-5.54	14.40	-0.02
	Neg Self-Mgmt Q2		-8.49	5.70	-0.09
	Neg Prep Q2		-6.80	6.10	-0.07
	Ambig Improv Q2		-19.42	6.28	-0.19*
Step 4	Ethnicity	.04*	27.35	3.54	0.45**
	Pos Acad Q1		12.72	8.00	0.10
	Neg Self-Mgmt Q1		-1.12	6.60	-0.01
	Neg Prep Q1		-21.08	7.98	-0.17**
	Ambig Improv Q1		-8.74	7.79	-0.07
	Pos Acad Q2		1.05	15.14	0.00
	Neg Self-Mgmt Q2		-9.02	5.64	-0.10
	Neg Prep Q2		-4.55	6.00	-0.04
	Ambig Improv Q2		-24.70	6.59	-0.24**
	Pos Acad Q3		14.18	6.15	0.14*
	Neg Self-Mgmt Q3		-11.81	8.31	-0.09
	Neg Prep Q3		-2.48	5.95	-0.02
	Ambig Improv Q3		13.39	5.95	0.14*
Step 5	Ethnicity	.18**	19.97	3.05	0.33**
	Pos Acad Q1		4.81	6.69	0.04
	Neg Self-Mgmt Q1		14.53	5.74	0.14*
	Neg Prep Q1		-7.38	6.79	-0.06
	Ambig Improv Q1		0.26	6.54	0.00
	Pos Acad Q2		10.98	12.61	0.05
	Neg Self-Mgmt Q2		4.21	4.90	0.05
	Neg Prep Q2		0.33	5.01	0.00
	Ambig Improv Q2		-12.98	5.62	-0.12*
	Pos Acad Q3		7.39	5.16	0.07
	Neg Self-Mgmt Q3		-7.99	6.91	-0.06
	Neg Prep Q3		1.22	4.96	0.01**
	Ambig Improv Q3		12.79	4.93	0.14**

	MA Letter Grades Q1		16.42	1.80	0.59
Step 6	Ethnicity	.02**	19.45	2.99	0.32**
	Pos Acad Q1		3.83	6.57	0.03
	Neg Self-Mgmt Q1		14.52	5.62	0.14*
	Neg Prep Q1		-9.69	6.70	-0.08
	Ambig Improv Q1		0.31	6.41	0.00
	Pos Acad Q2		9.73	12.37	0.04
	Neg Self-Mgmt Q2		9.35	5.11	0.10
	Neg Prep Q2		2.47	4.96	0.02
	Ambig Improv Q2		-11.62	5.52	-0.11*
	Pos Acad Q3		6.67	5.06	0.06
	Neg Self-Mgmt Q3		-6.35	6.79	-0.05
	Neg Prep Q3		1.88	4.86	0.02
	Ambig Improv Q3		15.44	4.92	0.16**
	MA Letter Grades Q1		11.34	2.49	0.41**
	MA Letter Grades Q2		6.43	2.22	0.26**
Step 7	Ethnicity	.01*	18.10	3.01	0.30**
	Pos Acad Q1		3.93	6.48	0.03
	Neg Self-Mgmt Q1		13.19	5.58	0.13*
	Neg Prep Q1		-6.49	6.74	-0.05
	Ambig Improv Q1		1.69	6.35	0.01
	Pos Acad Q2		8.67	12.21	0.04
	Neg Self-Mgmt Q2		10.49	5.07	0.11*
	Neg Prep Q2		5.01	5.01	0.05
	Ambig Improv Q2		-12.44	5.46	-0.12*
	Pos Acad Q3		10.50	5.24	0.10*
	Neg Self-Mgmt Q3		-9.24	6.81	-0.07
	Neg Prep Q3		-1.64	5.02	-0.02
	Ambig Improv Q3		14.68	4.87	0.16**
	MA Letter Grades Q1		12.44	2.50	0.45**
	MA Letter Grades Q2		9.45	2.52	0.38**
	MA Letter Grades Q3		-5.02	2.08	-0.20*

Note: R^2 = .64 (ΔR^2s add up to .65 due to rounding); n = 195. For ethnicity, 0 = Black, 1 = White. For comments, 0 = did not receive this comment, 1 = did receive this comment. Pos is positive. Neg is negative. Ambig is ambiguous. Atnd is the attendance comments. Self-Mgmt is the self-management comment. Other SEL is the other SEL comment. Prep is the preparation comment. Acad is the academics comment. Improv is the improvement comment. Academic quarters one, two, three and four are abbreviated as Q1, Q2, Q3, and Q4 respectively.

* $p < .05$, ** $p < .01$

Figure 1.
Partial Theoretical Model of How SEL Leads to Better Academic Performance (Inside School Only)

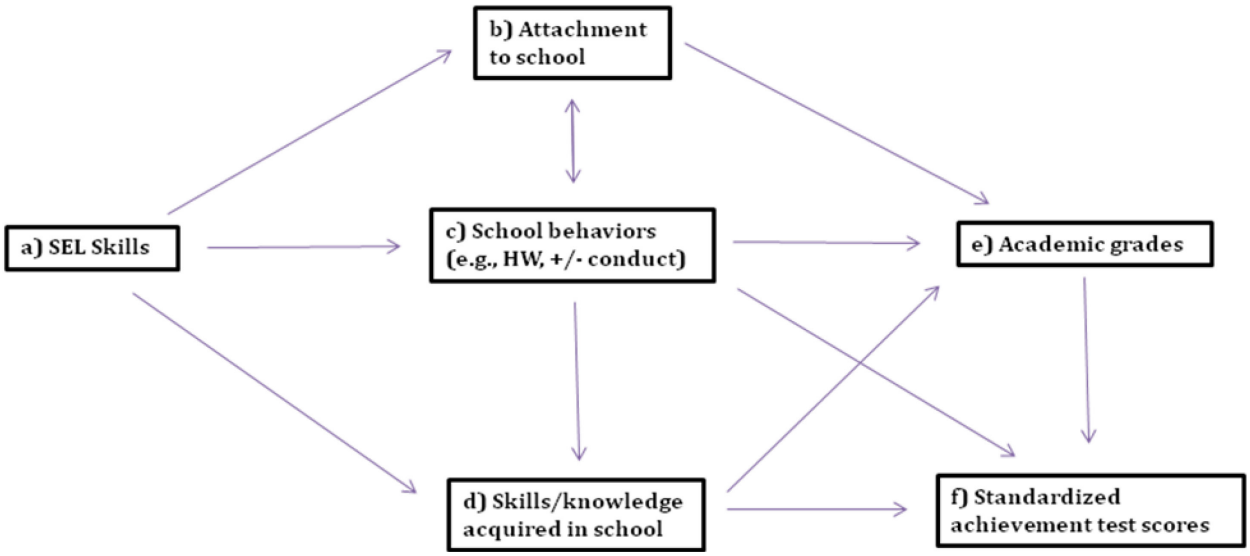


Figure 2.
Theoretical Model of How SEL and Intelligence Lead to Better Academic Performance

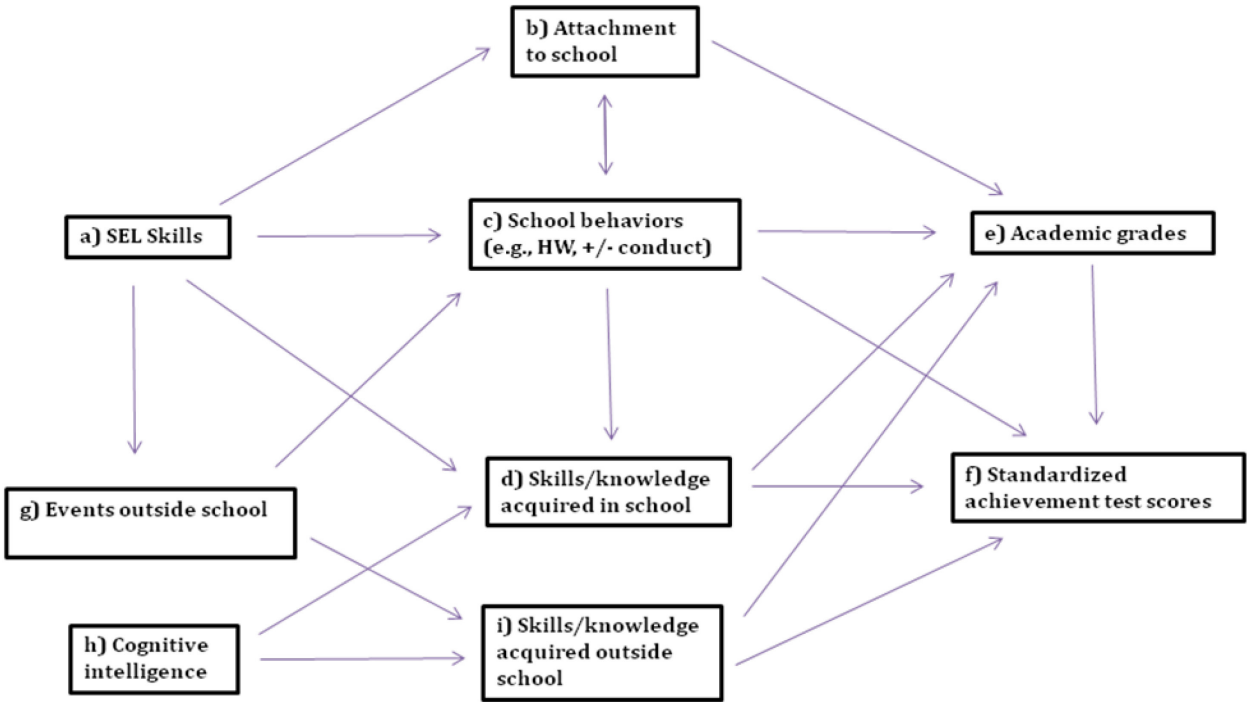


Figure 3.
Logic Model for the Components of the Report Card

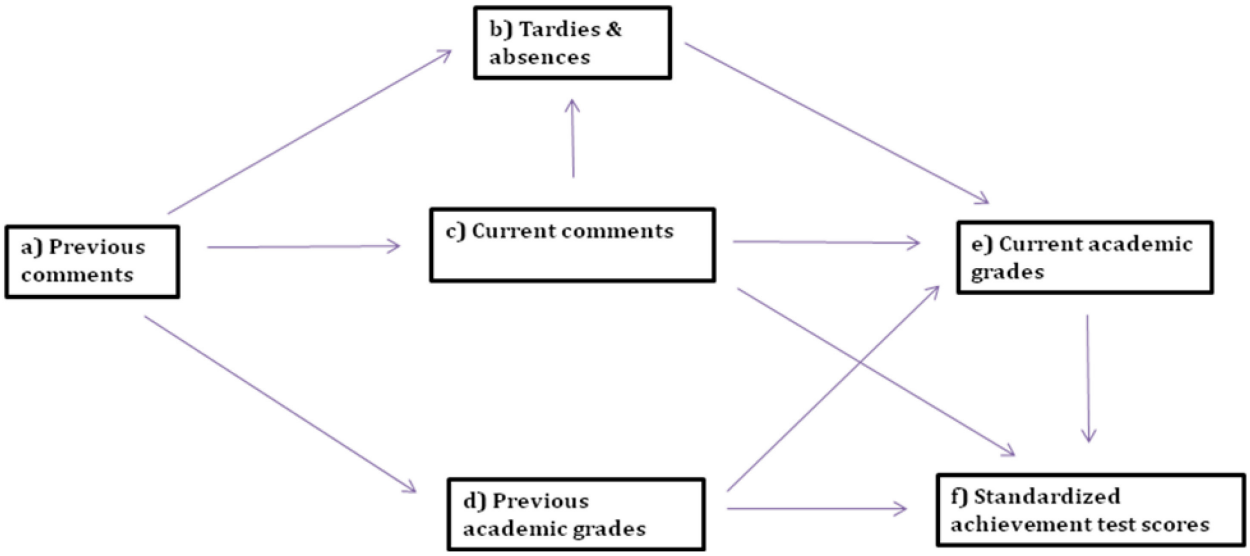
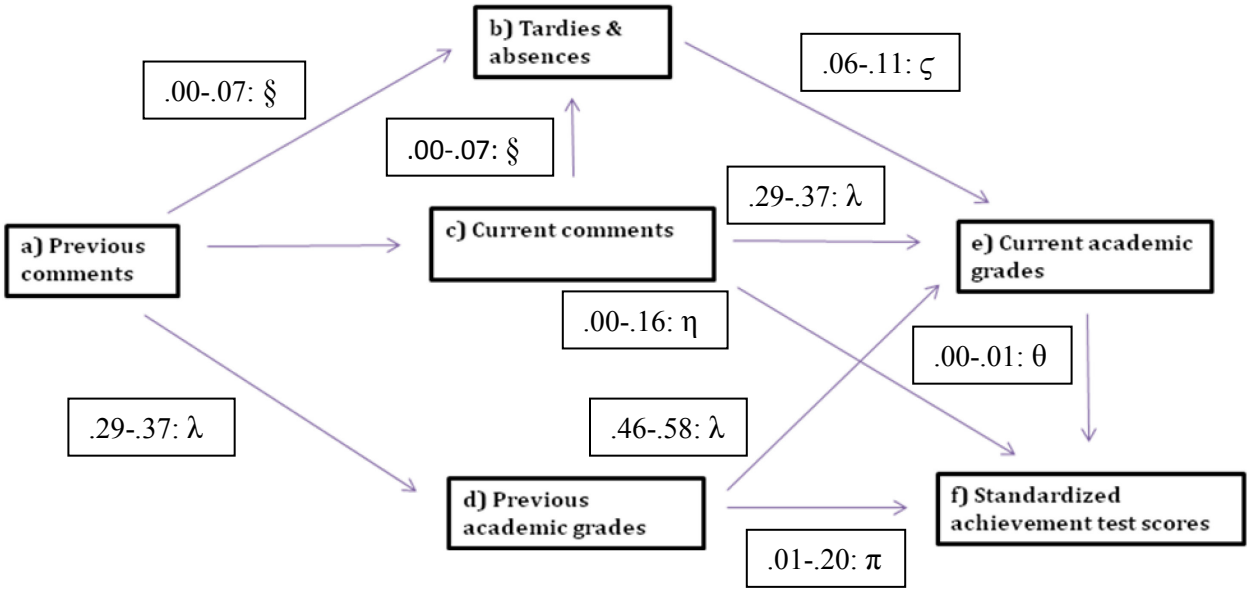


Figure 4.
Logic Model for the Components of the Report Card: Stengths of the Relationships



Note: The numbers in this figure were taken from the ΔR^2 values for the corresponding multiple regression steps for Q1 for language/literature and math in Tables 8 through 15, which are representative of the relationships in Q2 through Q4. The only exception is the values between the academic grades (i.e., d to e), where the bivariate correlations from the overall results were squared to get a rough equivalent to ΔR^2. There is no effect size provided for the two comments (i.e., a to c) because there is no equivalent ΔR^2 value for comparing them. As the ΔR^2 values come from different multiple regression analyses, the ΔR^2 steps occur at different parts of the multiple regression analyses, and the numbers are not a result of structural equation modeling (SEM), the figure should be read for the general principals expressed rather than the exact numbers.

θ = not significant to less than a small effect size, § = not significant to a small effect size, η = not significant to a medium effect size, ς = small effect size, π = small to medium effect size, λ = large effect size